

Form PTO 1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE (REV 5-93)		ATTORNEY'S DOCKET NUMBER B45110
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED / ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. APPLICATION NO. (If known, see 37 C.F.R. 1.5) <b>09/509239</b>
INTERNATIONAL APPLICATION NO. PCT/EP98/06040	INTERNATIONAL FILING DATE 17 September 1998	PRIORITY DATE CLAIMED 26 September 1997
TITLE OF INVENTION FUSION PROTEINS COMPRISING HIV-1 TAT AND/OR NEF PROTEINS		
APPLICANT(S) FOR DO/EO/US Claudine BRUCK, Stephane Andre Georges GODART and Martine MARC-HAND		

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
  - a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

**Items 11. to 16. below concern other document(s) or information included:**

11. ☒ An Information Disclosure Statement under 37 C.F.R. 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 C.F.R. 3.28 and 3.31 is included.
13. ☒ A **FIRST** preliminary amendment.  
☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.  
☒ Please amend the specification by inserting before the first line the sentence: This is a 371 of International Application PCT/EP98/06040, filed 17 September 1998, which claims benefit from the following Provisional Application, GB 9720585.0 filed 26 September 1997.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☐ Other items or information:

RECEIVED "6260560

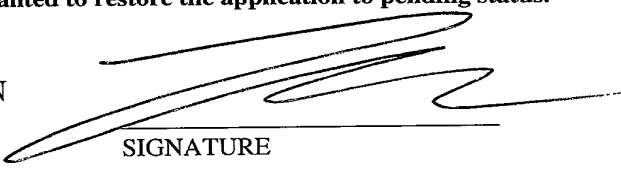
416 Rec'd PCT/PTO 23 MAR 2000

US APPLICATION NO. (if known see 37 CFR 1.50) <b>09/509239</b>		INTERNATIONAL APPLICATION NO. PCT/EP98/06040		ATTORNEYS DOCKET NO. B45110	
17. <input checked="" type="checkbox"/> The following fees are submitted:				CALCULATIONS PTO USE ONLY	
<b>Basic National Fee (37 C.F.R. 1.492(a)(1)-(5)):</b>					
Search Report has been prepared by the EPO or JPO .....\$840.00					
International Preliminary Examination Fee paid to USPTO (37 CFR 1.482) .....\$670.00					
No International Preliminary Examination Fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2)) .....\$690.00					
Neither International Preliminary Examination Fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....\$970.00					
International Preliminary Examination Fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4).....\$96.00					
<b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>				<b>\$840.00</b>	
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				<b>\$0.00</b>	
Claims	Number Filed	Number Extra	Rate		
Total claims	46 - 20 =	26	26 x \$18.00	<b>\$468.00</b>	
Independent claims	4 - 3 =	1	1 x \$78.00	<b>\$78.00</b>	
Multiple dependent claims (if applicable)			+ \$260.00	<b>\$260.00</b>	
<b>TOTAL OF ABOVE CALCULATIONS =</b>				<b>\$806.00</b>	
Reduction by 1/2 for filing by small entity, if applicable. Verified Small Entity statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28).				<b>\$</b>	
<b>SUBTOTAL =</b>				<b>\$1646.00</b>	
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)) +				<b>\$</b>	
<b>TOTAL NATIONAL FEE =</b>				<b>\$1646.00</b>	
				Amount to be refunded	\$
				charged	\$

- a. ☐ A check in the amount of \$\_\_\_\_\_ to cover the above fees is enclosed.
- b. ☒ Please charge my Deposit Account No. 19-2570 in the amount of **\$1646.00** to cover the above fees.  
A duplicate copy of this sheet is enclosed.
- c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 19-2570. A duplicate copy of this sheet is enclosed.
- d. ☒ General Authorization to charge any and all fees under 37 CFR 1.16 or 1.17, including petitions for extension of time relating to this application (37 CFR 1.136 (a)(3)).

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:  
SMITHKLINE BEECHAM CORPORATION  
Corporate Intellectual Property - UW2220  
P.O. Box 1539  
King of Prussia, PA 19406-0939  
Phone (610) 270-5024  
Facsimile (610) 270-5090

  
SIGNATURE  
Zoltan Kerekes  
NAME  
38,938  
REGISTRATION NO.

09/509239-06260560

416 Rec'd PCT/PTO 23 MAR 2000

32. A vaccine composition which comprises a protein comprising
- (a) an HIV Tat protein or derivative thereof linked to either (i) a fusion partner or (ii) an HIV Nef protein or derivative thereof; or
  - (b) an HIV Nef protein or derivative thereof linked to either (i) a fusion partner or (ii) an HIV Tat protein or derivative thereof; or
  - (c) an HIV Nef protein or derivative thereof linked to an HIV Tat protein or derivative thereof and a fusion partner,
- in admixture with a pharmaceutically acceptable excipient.
33. A composition as claimed in claim 32, comprising a Tat-Nef fusion protein or derivative thereof.

34. A composition as claimed in claim 32, comprising a Nef-Tat fusion protein or derivative thereof.
35. A composition according to claim 32, wherein the derivative of the Tat protein is a mutated Tat protein.
36. A composition according to claim 32, wherein the derivative of the Nef protein is a mutated Nef protein.
37. A composition as claimed in claim 32, wherein the fusion partner is a lipoprotein or derivative thereof.
38. A composition as claimed in claim 37, wherein the lipoprotein is Haemophilus Influenza B protein D or derivative thereof.
39. A composition as claimed in claim 38, wherein the fusion partner comprises between 100-130 amino acid from the N terminal of Haemophilus Influenza B protein D.
40. A composition as claimed in claim 32, wherein the Tat protein is the entire Tat protein.
41. A composition as claimed in claim 32, wherein the Nef protein is the entire Nef protein.
42. A composition as claimed in claim 32, wherein the Tat protein is fused to an HIV Nef protein and a fusion partner.
43. A composition as claimed in claim 32, wherein the protein has a Histidine tail.
44. A composition as claimed in claim 32, wherein the protein is carboxymethylated.
45. A composition as claimed in claim 32, additionally comprising an adjuvant.
46. A composition as claimed in claim 45, wherein the adjuvant is a TH1 inducing adjuvant.

47. A composition as claimed in claim 45 which adjuvant comprises monophosphoryl lipid A or a derivative thereof such as 3 de-O-acylated monophosphoryl lipid A.
48. A composition as claimed in claim 45, additionally comprising a saponin adjuvant.
49. A composition as claimed in any one of claims 45 to 48 which additionally comprises an oil in water emulsion.
50. A composition as claimed in claim 32 further comprising HIV gp160 or its derivative gp120.
51. A composition as claimed in claim 45 further comprising HIV gp160 or its derivative gp120.
52. A composition as claimed in claim 48 further comprising HIV gp160 or its derivative gp120.
53. A composition as claimed in claim 49 further comprising HIV gp160 or its derivative gp120.
54. A protein comprising an HIV Tat protein or derivative thereof linked to an HIV Nef protein or derivative thereof in Nef-Tat or Tat-Nef orientation.
55. A nucleic acid encoding a protein of claim 54.
56. A host transformed with a nucleic acid of claim 55.
57. A host as claimed in claim 56 wherein the host is either *E. coli* or *Pichia pastoris*.
58. A method of producing a protein of claim 54, comprising providing a host as claimed in claim 56 or 57, expressing said protein and recovering the protein.

59. A method of preparing (i) an HIV Nef protein or derivative thereof or (ii) an HIV Tat protein or derivative thereof in *Pichia pastoris* which method comprises the steps of transforming *Pichia pastoris* with DNA encoding said HIV Nef protein or derivative thereof of HIV Tat protein or derivative thereof, expressing said protein and recovering the protein.
60. The method of claim 58 further comprising a carboxymethylation step performed on the expressed protein.
61. The method of claim 59 further comprising a carboxymethylation step performed on the expressed protein.
62. A method of producing a vaccine, comprising admixing the protein from claim 58 with a pharmaceutically acceptable diluent.
63. A method of producing a vaccine, comprising admixing the protein from claim 59 with a pharmaceutically acceptable diluent.
64. A method of producing a vaccine, comprising admixing the protein from claim 60 with a pharmaceutically acceptable diluent.
65. The method of claim 62 further comprising the addition of HIV gp160 or its derivative gp120.
66. The method of claim 63 further comprising the addition of HIV gp160 or its derivative gp120.
67. The method of claim 64 further comprising the addition of HIV gp160 or its derivative gp120.



Intl. App. No.: PCT/EP98/06040  
Docket No. B45110

## REMARKS

The above-identified application is being entered into the National Phase from PCT application no. PCT/EP98/06040.

Applicants have deleted claims 1-31 and added new claims 32-77 to put the claims in conformity with U.S. practice.

No new matter has been introduced.

Respectfully submitted,



Zoltan Kerekes  
Attorney for Applicants  
Registration No. 38,938

SMITHKLINE BEECHAM CORPORATION  
Corporate Intellectual Property - UW2220  
P.O. Box 1539  
King of Prussia, PA 19406-0939  
Phone (610) 270-5024  
Facsimile (610) 270-5090

N:\zk\apps\B45110\PREAMD.DOC



## FUSION PROTEINS COMPRISING HIV-1 TAT AND/OR NEF PROTEINS

- 1 -

third, i.e. approximately the first 100-130 amino acids are utilised. This is represented herein as Lipo D 1/3. In a preferred embodiment of the invention the Nef protein or derivative thereof may be linked to the Tat protein or derivative thereof. Such Nef-Tat fusions may optionally also be linked to an fusion partner, such as protein D.

5

The fusion partner is normally linked to the N-terminus of the Nef or Tat protein.

Derivatives encompassed within the present invention include molecules with a C terminal Histidine tail which preferably comprises between 5-10 Histidine residues.

10 Generally, a histidine tail containing n residues is represented herein as His (n). The presence of an histidine (or 'His') tail aids purification. More specifically, the invention provides proteins with the following structure

15	Lipo D 1/3	-	Nef	-	His (6)
	Lipo D 1/3	-	Nef-Tat	-	His (6)
	Prot D 1/3	-	Nef	-	His (6)
20	Prot D 1/3	-	Nef-Tat	-	His (6)
			Nef-Tat	-	His (6)

Figure 1 provides the amino-acid (Seq. ID. No. 7) and DNA sequence (Seq. ID. No. 6) of the fusion partner for such constructs.

In a preferred embodiment the proteins are expressed with a Histidine tail comprising between 5 to 10 and preferably six Histidine residues. These are advantageous in aiding purification. Separate expression, in yeast (*Saccharomyces cerevisiae*), of Nef (Macreadie I.G. et al., 1993, *Yeast* 9 (6) 565-573) and Tat (Braddock M et al., 1989, *Cell* 58 (2) 269-79) has already been reported. Nef protein only is myristilated. The present invention provides for the first time the expression of Nef and Tat separately

in a Pichia expression system (Nef-His and Tat-His constructs), and the successful expression of a fusion construct Nef-Tat-His. The DNA and amino acid sequences of representative Nef-His (Seq. ID. No.s 8 and 9), Tat-His (Seq. ID. No.s 10 and 11) and of Nef-Tat-His fusion proteins (Seq. ID. No.s 12 and 13) are set forth in Figure 2.

5

Derivatives encompassed within the present invention also include mutated proteins. The term 'mutated' is used herein to mean a molecule which has undergone deletion, addition or substitution of one or more amino acids using well known techniques for site directed mutagenesis or any other conventional method.

10

A mutated Tat is illustrated in Figure 2 (Seq. ID. No.s 22 and 23) as is a Nef-Tat Mutant-His (Seq. ID. No.s 24 and 25).

15

The present invention also provides a DNA encoding the proteins of the present invention. Such sequences can be inserted into a suitable expression vector and expressed in a suitable host.

20

A DNA sequence encoding the proteins of the present invention can be synthesized using standard DNA synthesis techniques, such as by enzymatic ligation as described by D.M. Roberts *et al.* in Biochemistry 1985, 24, 5090-5098, by chemical synthesis, by *in vitro* enzymatic polymerization, or by PCR technology utilising for example a heat stable polymerase, or by a combination of these techniques.

25

Enzymatic polymerisation of DNA may be carried out *in vitro* using a DNA polymerase such as DNA polymerase I (Klenow fragment) in an appropriate buffer containing the nucleoside triphosphates dATP, dCTP, dGTP and dTTP as required at a temperature of 10°-37°C, generally in a volume of 50µl or less. Enzymatic ligation of DNA fragments may be carried out using a DNA ligase such as T4 DNA ligase in an appropriate buffer, such as 0.05M Tris (pH 7.4), 0.01M MgCl<sub>2</sub>, 0.01M dithiothreitol, 1mM spermidine, 1mM ATP and 0.1mg/ml bovine serum albumin, at a temperature of 4°C to ambient, generally in a volume of 50ml or less. The chemical synthesis of the DNA polymer or fragments may be carried out by conventional

30

- phosphotriester, phosphite or phosphoramidite chemistry, using solid phase techniques such as those described in 'Chemical and Enzymatic Synthesis of Gene Fragments - A Laboratory Manual' (ed. H.G. Gassen and A. Lang), Verlag Chemie, Weinheim (1982), or in other scientific publications, for example M.J. Gait, H.W.D. Matthes, M. Singh, B.S. Sproat, and R.C. Titmas, *Nucleic Acids Research*, 1982, 10, 6243; B.S. Sproat, and W. Bannwarth, *Tetrahedron Letters*, 1983, 24, 5771; M.D. Matteucci and M.H. Caruthers, *Tetrahedron Letters*, 1980, 21, 719; M.D. Matteucci and M.H. Caruthers, *Journal of the American Chemical Society*, 1981, 103, 3185; S.P. Adams *et al.*, *Journal of the American Chemical Society*, 1983, 105, 661; N.D. Sinha, J. Biernat, J. McMannus, and H. Koester, *Nucleic Acids Research*, 1984, 12, 4539; and H.W.D. Matthes *et al.*, *EMBO Journal*, 1984, 3, 801.

The invention also provides a process for preparing a protein of the invention, the process comprising the steps of :

- i) preparing a replicable or integrating expression vector capable, in a host cell, of expressing a DNA polymer comprising a nucleotide sequence that encodes the protein or a derivative thereof
- ii) transforming a host cell with said vector
- iii) culturing said transformed host cell under conditions permitting expression of said DNA polymer to produce said protein; and
- iv) recovering said protein

The process of the invention may be performed by conventional recombinant techniques such as described in Maniatis *et al.*, *Molecular Cloning - A Laboratory Manual*; Cold Spring Harbor, 1982-1989.

The term 'transforming' is used herein to mean the introduction of foreign DNA into a host cell. This can be achieved for example by transformation, transfection or

infection with an appropriate plasmid or viral vector using e.g. conventional techniques as described in Genetic Engineering; Eds. S.M. Kingsman and A.J. Kingsman; Blackwell Scientific Publications; Oxford, England, 1988. The term 'transformed' or 'transformant' will hereafter apply to the resulting host cell  
5 containing and expressing the foreign gene of interest.

The expression vectors are novel and also form part of the invention.

The replicable expression vectors may be prepared in accordance with the invention,  
10 by cleaving a vector compatible with the host cell to provide a linear DNA segment having an intact replicon, and combining said linear segment with one or more DNA molecules which, together with said linear segment encode the desired product, such as the DNA polymer encoding the protein of the invention, or derivative thereof, under ligating conditions.

15 Thus, the DNA polymer may be preformed or formed during the construction of the vector, as desired.

The choice of vector will be determined in part by the host cell, which may be  
20 prokaryotic or eukaryotic but preferably is *E. coli* or yeast. Suitable vectors include plasmids, bacteriophages, cosmids and recombinant viruses.

The preparation of the replicable expression vector may be carried out conventionally with appropriate enzymes for restriction, polymerisation and ligation of the DNA, by  
25 procedures described in, for example, Maniatis *et al.* cited above.

The recombinant host cell is prepared, in accordance with the invention, by transforming a host cell with a replicable expression vector of the invention under transforming conditions. Suitable transforming conditions are conventional and are  
30 described in, for example, Maniatis *et al.* cited above, or "DNA Cloning" Vol. II, D.M. Glover ed., IRL Press Ltd, 1985.

- The choice of transforming conditions is determined by the host cell. Thus, a bacterial host such as *E. coli* may be treated with a solution of  $\text{CaCl}_2$  (Cohen *et al.*, Proc. Nat. Acad. Sci., 1973, 69, 2110) or with a solution comprising a mixture of  $\text{RbCl}$ ,  $\text{MnCl}_2$ , potassium acetate and glycerol, and then with 3-[N-morpholino]-propane-sulphonic acid,  $\text{RbCl}$  and glycerol. Mammalian cells in culture may be transformed by calcium co-precipitation of the vector DNA onto the cells. The invention also extends to a host cell transformed with a replicable expression vector of the invention.
- 10 Culturing the transformed host cell under conditions permitting expression of the DNA polymer is carried out conventionally, as described in, for example, Maniatis *et al.* and "DNA Cloning" cited above. Thus, preferably the cell is supplied with nutrient and cultured at a temperature below 50°C.
- 15 The product is recovered by conventional methods according to the host cell. Thus, where the host cell is bacterial, such as *E. coli* - or yeast such as *Pichia*; it may be lysed physically, chemically or enzymatically and the protein product isolated from the resulting lysate. Where the host cell is mammalian, the product may generally be isolated from the nutrient medium or from cell free extracts. Conventional protein
- 20 isolation techniques include selective precipitation, adsorption chromatography, and affinity chromatography including a monoclonal antibody affinity column.

- For proteins of the present invention provided with Histidine tails, purification can easily be achieved by the use of a metal ion affinity column. In a preferred
- 25 embodiment, the protein is further purified by subjecting it to cation ion exchange chromatography and/or Gel filtration chromatography. The protein is then sterilised by passing through a 0.22  $\mu\text{m}$  membrane.

- The proteins of the invention can then be formulated as a vaccine, or the Histidine
- 30 residues enzymatically cleared.

The proteins of the present invention are provided preferably at least 80% pure more preferably 90% pure as visualised by SDS PAGE. Preferably the proteins appear as a single band by SDS PAGE.

- 5 The present invention also provides pharmaceutical composition comprising a protein of the present invention in a pharmaceutically acceptable excipient.

Vaccine preparation is generally described in **New Trends and Developments in Vaccines**, Voller *et al.* (eds.), University Park Press, Baltimore, Maryland, 1978.

- 10 Encapsulation within liposomes is described by Fullerton, US Patent 4,235,877.

- The proteins of the present invention are preferably adjuvanted in the vaccine formulation of the invention. Suitable adjuvants include an aluminium salt such as aluminium hydroxide gel (alum) or aluminium phosphate, but may also be a salt of calcium, iron or zinc, or may be an insoluble suspension of acylated tyrosine, or  
15 acylated sugars, cationically or anionically derivatised polysaccharides, or polyphosphazenes.

- In the formulation of the inventions it is preferred that the adjuvant composition  
20 induces a preferential TH1 response. Suitable adjuvant systems include, for example, a combination of monophosphoryl lipid A or derivative thereof, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL) together with an aluminium salt.

- An enhanced system involves the combination of a monophosphoryl lipid A and a  
25 saponin derivative particularly the combination of QS21 and 3D-MPL as disclosed in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol as disclosed in WO 96/33739.

- A particularly potent adjuvant formulation involving QS21, 3D-MPL & tocopherol in  
30 an oil in water emulsion is described in WO 95/17210 and is a preferred formulation.

Accordingly in one embodiment of the present invention there is provided a vaccine comprising a protein according to the invention adjuvanted with a monophosphoryl lipid A or derivative thereof, especially 3D-MPL.

- 5 Preferably the vaccine additionally comprises a saponin, more preferably QS21.

Preferably the formulation additionally comprises an oil in water emulsion and tocopherol. The present invention also provides a method for producing a vaccine formulation comprising mixing a protein of the present invention together with a  
10 pharmaceutically acceptable excipient, such as 3D-MPL.

The vaccine of the present invention may additionally comprise further HIV proteins, such as the envelope glycoprotein gp160 or its derivative gp 120.

- 15 In another aspect, the invention relates to an HIV Nef or an HIV Tat protein or derivative thereof expressed in *Pichia pastoris*.

The invention will be further described by reference to the following examples:

20 **EXAMPLES:**

**General**

Nef and Tat proteins, two regulatory proteins encoded by the human  
25 immunodeficiency virus (HIV-1) were produced in *E.coli* and in the methylotrophic yeast *Pichia pastoris*.

The *nef* gene from the Bru/Lai isolate (Cell 40: 9-17, 1985) was selected for these constructs since this gene is among those that are most closely related to the  
30 consensus Nef.



The starting material for the Bru/Lai *nef* gene was a 1170bp DNA fragment cloned on the mammalian expression vector pcDNA3 (pcDNA3/*nef*).

The *tat* gene originates from the BH10 molecular clone. This gene was received as an  
5 HTLV III cDNA clone named pCV1 and described in Science, 229, p69-73, 1985.

# 1. EXPRESSION OF HIV-1 *nef* AND *tat* SEQUENCES IN E.COLI.

Sequences encoding the Nef protein as well as a fusion of *nef* and *tat* sequences were  
10 placed in plasmids vectors: pRIT14586 and pRIT14589 (see figure 1).

Nef and the Nef-Tat fusion were produced as fusion proteins using as fusion partner a  
part of the protein D. Protein D is an immunoglobulin D binding protein exposed at  
the surface of the gram-negative bacterium *Haemophilus influenzae*.  
15

pRIT14586 contains, under the control of a  $\lambda$ PL promoter, a DNA sequence derived  
from the bacterium *Haemophilus influenzae* which codes for the first 127 amino acids  
of the protein D (Infect. Immun. 60 : 1336-1342, 1992), immediately followed by a  
multiple cloning site region plus a DNA sequence coding for one glycine, 6 histidines  
20 residues and a stop codon (Fig. 1A).

This vector is designed to express a processed lipidated His tailed fusion protein  
(LipoD fusion protein). The fusion protein is synthesised as a precursor with an 18  
amino acid residues long signal sequence and after processing, the cysteine at position  
25 19 in the precursor molecule becomes the amino terminal residue which is then  
modified by covalently bound fatty acids (Fig.1B).

pRIT14589 is almost identical to pRIT14586 except that the protD derived sequence  
starts immediately after the cysteine19 codon.

30 Expression from this vector results in a His tailed, non lipidated fusion protein  
(Prot D fusion protein).

## 10

## 15

15

PRIMER 01 (Seq ID NO 1): 5'ATCGTCCATG.GGT.GGC.AAG.TGG.T 3'

20

25

25

30 The PCR fragment obtained and the expression plasmid pRIT14586 were both  
restricted by NcoI and SpeI, purified on an agarose gel, ligated and transformed in the

appropriate *E. coli* host cell, strain AR58. This strain is a cryptic  $\lambda$  lysogen derived from N99 that is *galE::Tn10*,  $\Delta$ -8 (*chlD-pgl*),  $\Delta$ -H1 (*cro-chlA*),  $N^+$ , and *cl857*.

The resulting recombinant plasmid received, after verification of the *nef* amplified  
5 region by automatic sequencing, (see section 1.1.2 below) the pRIT14595  
denomination.

### 1.1.2 Selection of transformants of *E. Coli* strain AR58 with pRIT14595

10

When transformed in AR58 *E. coli* host strain, the recombinant plasmid directs the heat-inducible production of the heterologous protein.

Heat inducible protein production of several recombinant lipoD-Nef-His  
15 transformants was analysed by Coomassie Blue stained SDS-PAGE. All the transformants analysed showed an heat inducible heterologous protein production. The abundance of the recombinant Lipo D-Nef-Tat-His fusion protein was estimated at 10% of total protein.

20 One of the transformants was selected and given the laboratory accession number ECLD-N1.

The recombinant plasmid was reisolated from strain ECLD-N1, and the sequence of the *nef*-His coding region was confirmed by automated sequencing. This plasmid  
25 received the official designation pRIT14595.

The fully processed and acylated recombinant Lipo D-*nef*-His fusion protein produced by strain ECLD-N1 is composed of:

30

°Fatty acids

°109 a.a. of proteinD (starting at a.a.19 and extending to a.a.127).

°A methionine, created by the use of NcoI cloning site of pRIT14586 (Fig.1).

°205a.a. of Nef protein (starting at a.a.2 and extending to a.a.206).

5 °A threonine and a serine created by the cloning procedure (cloning at SpeI site of pRIT14586).

°One glycine and six histidines.

## 1.2 CONSTRUCTION OF RECOMBINANT STRAIN ECD-N1 PRODUCING PROT D-Nef-HIS FUSION PROTEIN.

10

Construction of expression plasmid pRIT14600 encoding the Prot D-Nef-His fusion protein was identical to the plasmid construction described in example 1.1.1 with the exception that pRIT14589 was used as receptor plasmid for the PCR amplified *nef* fragment.

15

E.coli AR58 strain was transformed with pRIT14600 and transformants were analysed as described in example 1.1.2. The transformant selected received laboratory accession number ECD-N1.

### 1.3 CONSTRUCTION OF RECOMBINANT STRAIN ECLD-NT6 PRODUCING THE LIPO D-Nef-Tat-HIS FUSION PROTEIN.

#### 1.3.1 Construction of the lipo D-Nef-Tat-His expression plasmid pRIT14596

5

The *tat* gene(BH10 isolate) was amplified by PCR from a derivative of the pCV1 plasmid with primers 03 and 04. SpeI restriction sites were introduced at both ends of the PCR fragment.

10

SpeI

PRIMER 03 (Seq ID NO 3): 5' ATCGTACTAGT.GAG.CCA.GTA.GAT.C 3'

SpeI

PRIMER 04 (Seq ID NO 4): 5' CGGCTACTAGTTTTCCTTCGGGCCT 3'

15

The nucleotide sequence of the amplified *tat* gene is illustrated in the pCV1 clone (Science 229 : 69-73, 1985) and covers nucleotide 5414 till nucleotide 7998.

20

The PCR fragment obtained and the plasmid pRIT14595 (expressing lipoD-Nef-His protein) were both digested by SpeI restriction enzyme, purified on an agarose gel, ligated and transformed in competent AR58 cells. The resulting recombinant plasmid received, after verification of the *tat* amplified sequence by automatic sequencing (see section 1.3.2 below), the pRIT14596 denomination.

25

#### 1.3.2 Selection of transformants of strain AR58 with pRIT14596

Transformants were grown, heat induced and their proteins were analysed by Coomassie Blue stained gels. The production level of the recombinant protein was estimated at 1% of total protein. One recombinant strain was selected and received the laboratory denomination ECLD-NT6.

30

The lipoD-*nef-tat*-His recombinant plasmid was reisolated from ECLD-NT6 strain, sequenced and received the official designation pRIT14596.

The fully processed and acylated recombinant Lipo D-Nef-Tat-His fusion protein  
5 produced by strain ECLD-N6 is composed of:

°Fatty acids

°109 a.a. of proteinD (starting at a.a.19 and extending to a.a.127).

°A methionine, created by the use of NcoI cloning site of pRIT14586.

10 °205a.a. of the Nef protein (starting at a.a.2 and extending to a.a.206)

°A threonine and a serine created by the cloning procedure

°85a.a. of the Tat protein (starting at a.a.2 and extending to a.a.86)

°A threonine and a serine introduced by cloning procedure

°One glycine and six histidines.

15

#### 1.4 CONSTRUCTION OF RECOMBINANT STRAIN ECD-NT1 PRODUCING PROT D-Nef-Tat-HIS FUSION PROTEIN.

Construction of expression plasmid pRIT14601 encoding the Prot D-Nef-Tat-His  
20 fusion protein was identical to the plasmid construction described in example 1.3.1  
with the exception that pRIT14600 was used as receptor plasmid for the PCR  
amplified *nef* fragment.

*E.coli* AR58 strain was transformed with pRIT14601 and transformants were analysed  
25 as described previously. The transformant selected received laboratory accession  
number ECD-NT1.

30

## 2. EXPRESSION OF HIV-1 *nef* AND *tat* SEQUENCES IN *PICHIA PASTORIS*.

Nef protein, Tat protein and the fusion Nef -Tat were expressed in the methylotrophic yeast *Pichia pastoris* under the control of the inducible alcohol oxidase (AOX1) promoter.

To express these HIV-1 genes a modified version of the integrative vector PHIL-D2 (INVITROGEN) was used. This vector was modified in such a way that expression of heterologous protein starts immediately after the native ATG codon of the AOX1 gene and will produce recombinant protein with a tail of one glycine and six histidines residues. This PHIL-D2-MOD vector was constructed by cloning an oligonucleotide linker between the adjacent *Asu*II and *Eco*RI sites of PHIL-D2 vector (see Figure 3). In addition to the His tail, this linker carries *Nco*I, *Spe*I and *Xba*I restriction sites between which *nef*, *tat* and *nef-tat* fusion were inserted.

### 2.1 CONSTRUCTION OF THE INTEGRATIVE VECTORS pRIT14597 (encoding Nef-His protein), pRIT14598 (encoding Tat-His protein) and pRIT14599 (encoding fusion Nef-Tat-His).

The *nef* gene was amplified by PCR from the pcDNA3/Nef plasmid with primers 01 and 02(see section 1.1.1 construction of pRIT14595).The PCR fragment obtained and the integrative PHIL-D2-MOD vector were both restricted by *Nco*I and *Spe*I, purified on agarose gel and ligated to create the integrative plasmid pRIT14597 (see Figure 3).

The *tat* gene was amplified by PCR from a derivative of the pCV1 plasmid with primers 05 and 04(see section 1.3.1 construction of pRIT14596):

*Nco*I

PRIMER 05 (Seq ID NO 5): 5'ATCGTCCATGGAGCCAGTAGATC 3'

10

## 2.2 TRANSFORMATION OF PICHIA PASTORIS STRAIN GS115(his4)

15

20

25

30

- 16 -



<sup>a</sup>A threonine and a serine created by the cloning procedure (cloning at SpeI site of PHIL-D2-MOD vector).

<sup>o</sup>One glycine and six histidines.

5 Strain Y1739 (Mut<sup>+</sup> phenotype) producing the Tat-His protein, a 95 amino acid protein which is composed of:

<sup>o</sup>A methionine created by the use of NcoI cloning site

<sup>o</sup>85 a.a. of the Tat protein(starting at a.a.2 and extending to a.a.86)

10

<sup>o</sup>A threonine and a serine introduced by cloning procedure

<sup>o</sup>One glycine and six histidines

Strain Y1737(Mut<sup>s</sup> phenotype) producing the recombinant Nef-Tat-His fusion protein,  
15 a myristylated 302 amino acids protein which is composed of:

°Myristic acid

<sup>o</sup>A methionine, created by the use of NcoI cloning site

°205a.a. of Nef protein(starting at a.a.2 and extending to a.a.206)

20

<sup>o</sup>A threonine and a serine created by the cloning procedure

°85a.a. of the Tat protein(starting at a.a.2 and extending to a.a.86)

<sup>o</sup>A threonine and a serine introduced by the cloning procedure

<sup>o</sup>One glycine and six histidines

### 3. EXPRESSION OF HIV-1 Tat-MUTANT IN PICHIA PASTÓRIS

As well as a Nef-Tat mutant fusion protein, a mutant recombinant Tat protein has also  
5 been expressed. The mutant Tat protein must be **biologically inactive** while  
**maintaining its immunogenic epitopes.**

A double mutant *tat* gene, constructed by D.Clements (Tulane University) was  
selected for these constructs.

10

This *tat* gene (originates from BH10 molecular clone) bears **mutations** in the **active**  
**site region (Lys41→Ala)** and in **RGD motif (Arg78→Lys and Asp80→Glu)** (  
Virology 235: 48-64, 1997).

15 The mutant *tat* gene was received as a cDNA fragment subcloned between the EcoRI  
and HindIII sites within a CMV expression plasmid (pCMVLys41/KGE)

#### 3.1 CONSTRUCTION OF THE INTEGRATIVE VECTORS

20 **pRIT14912(encoding Tat mutant-His protein) and pRIT14913(encoding fusion**  
**Nef-Tat mutant-His).**

The *tat* mutant gene was amplified by PCR from the pCMVLys41/KGE plasmid with  
primers 05 and 04 (see section 2.1 construction of pRIT14598)

25

An NcoI restriction site was introduced at the 5' end of the PCR fragment while a  
SpeI site was introduced at the 3' end with primer 04. The PCR fragment obtained and  
the PHIL-D2-MOD vector were both restricted by NcoI and SpeI, purified on agarose  
gel and ligated to create the integrative plasmid pRIT14912

30

To construct pRIT14913, the *tat* mutant gene was amplified by PCR from the pCMVLys41/KGE plasmid with primers 03 and 04 (see section 1.3.1 construction of pRIT14596).

- 5 The PCR fragment obtained and the plasmid pRIT14597 (expressing Nef-His protein) were both digested by SpeI restriction enzyme, purified on agarose gel and ligated to create the integrative plasmid pRIT14913

### 3.2 TRANSFORMATION OF PICHIA PASTORIS STRAIN GS115.

10

Pichia pastoris strains expressing Tat mutant-His protein and the fusion Nef-Tat mutant-His were obtained, by applying integration and recombinant strain selection strategies previously described in section 2.2 .

- 15 Two recombinant strains producing Tat mutant-His protein ,a 95 amino-acids protein, were selected: Y1775 (Mut<sup>+</sup> phenotype) and Y1776(Mut<sup>+</sup> phenotype).

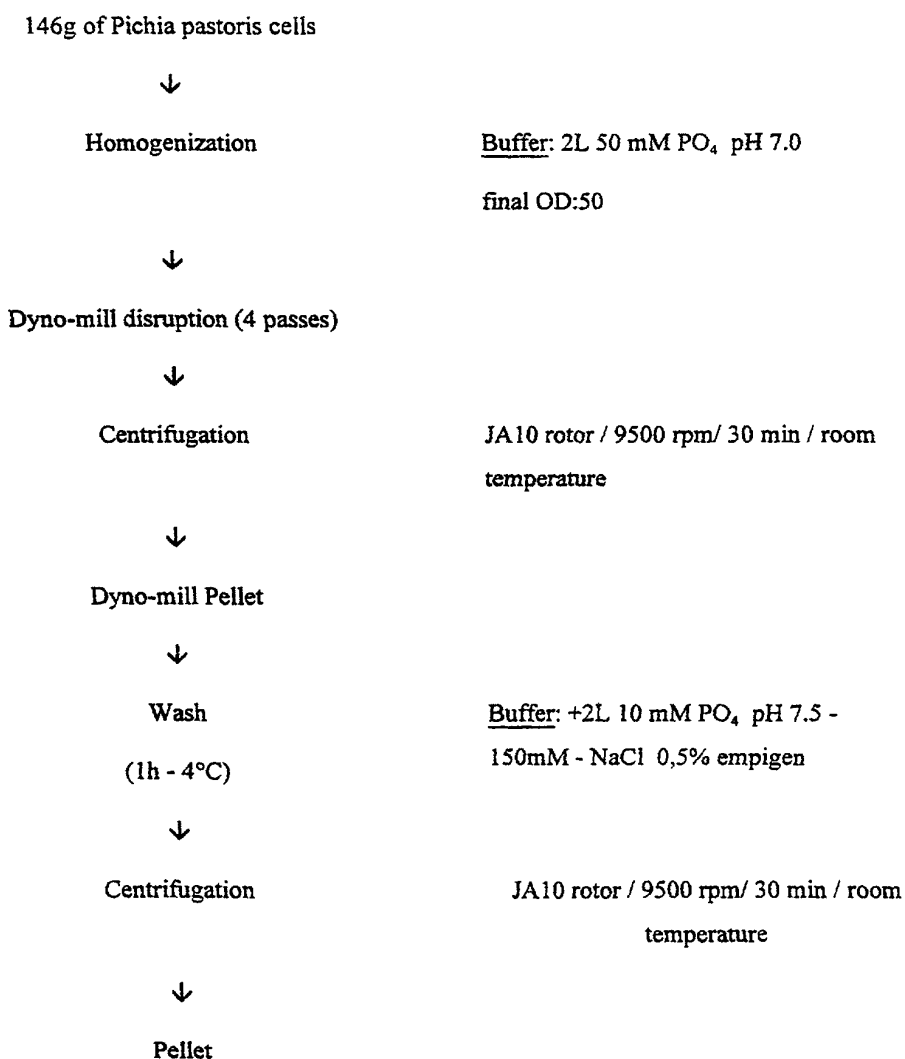
One recombinant strain expressing Nef-Tat mutant-His fusion protein, a 302 amino-acids protein was selected: Y1774(Mut<sup>+</sup> phenotype).

20

#### 4. PURIFICATION OF Nef-Tat-His FUSION PROTEIN (PICHIA PASTORIS)

- 5 The purification scheme has been developed from 146g of recombinant Pichia pastoris cells (wet weight) or 2L Dyno-mill homogenate OD 55. The chromatographic steps are performed at room temperature. Between steps, Nef-Tat positive fractions are kept overnight in the cold room (+4°C); for longer time, samples are frozen at -20°C.

10



↓	
Solubilisation (O/N - 4°C)	<u>Buffer</u> : + 660ml 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 4.0M GuHCl
↓	
Reduction (4H - room temperature - in the dark)	+ 0,2M 2-mercaptoethanesulfonic acid, sodium salt (powder addition) / pH adjusted to 7.5 (with 0,5M NaOH solution) before incubation
↓	
Carboxymethylation (1/2 h - room temperature - in the dark)	+ 0,25M Iodoacetamid (powder addition) / pH adjusted to 7.5 (with 0,5M NaOH solution) before incubation
↓	
Immobilized metal ion affinity chromatography on Ni <sup>2+</sup> -NTA-Agarose (Qiagen - 30 ml of resin)	<u>Equilibration buffer</u> : 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 4.0M GuHCl  <u>Washing buffer</u> : 1) Equilibration buffer  2) 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 6M Urea  3) 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 6M Urea - 25 mM Imidazol  <u>Elution buffer</u> : 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 6M Urea - 0,5M Imidazol
↓	
Dilution	Down to an ionic strength of 18 mS/cm <sup>2</sup>  <u>Dilution buffer</u> : 10 mM PO <sub>4</sub> pH 7.5 - 6M Urea
↓	
Cation exchange chromatography on SP Sepharose FF (Pharmacia - 30 ml of resin)	<u>Equilibration buffer</u> : 10 mM PO <sub>4</sub> pH 7.5 - 150mM NaCl - 6.0M Urea

Washing buffer: 1) Equilibration  
buffer

2) 10 mM PO<sub>4</sub> pH  
7.5 - 250mM NaCl - 6M Urea

Elution buffer: 10 mM Borate pH 9.0 -  
2M NaCl - 6M Urea

↓

Concentration

up to 5 mg/ml

10kDa Omega membrane(Filtron)

↓

Gel filtration chromatography on Superdex200 XK  
16/60

Elution buffer: 10 mM PO<sub>4</sub> pH 7.5 -  
150mM NaCl - 6M Urea

(Pharmacia - 120 ml of resin)

5 ml of sample / injection → 5 injections

↓

Dialysis

Buffer: 10 mM PO<sub>4</sub> pH 6.8 - 150mM

(O/N - 4°C)

NaCl - 0,5M Arginin\*

↓

Sterile filtration

Millex GV 0,22µm

\* ratio: 0,5M Arginin for a protein concentration of 1600µg/ml.

## 5 Purity

The level of purity as estimated by SDS-PAGE is shown in Figure 4 by Daiichi Silver Staining and in Figure 5 by Coomassie blue G250.

51mg of Nef-Tat-his protein are purified from 146g of recombinant *Pichia pastoris* cells (= 2L of Dyno-mill homogenate OD 55)

## 10 5. VACCINE PREPARATION

A vaccine prepared in accordance with the invention comprises the expression product of a DNA recombinant encoding an antigen as exemplified in example 1 or 2 and as adjuvant, the formulation comprising a mixture of 3 de -O-acylated monophosphoryl lipid A 3D-MPL and QS21 in an oil/water emulsion.

**3D-MPL:** is a chemically detoxified form of the lipopolysaccharide (LPS) of the Gram-negative bacteria *Salmonella minnesota*.

20 Experiments performed at Smith Kline Beecham Biologicals have shown that  
3D-MPL combined with various vehicles strongly enhances both the humoral and a  
TH1 type of cellular immunity.

**QS21:** is one saponin purified from a crude extract of the bark of the *Quillaja*

25 Saponaria Molina tree, which has a strong adjuvant activity: it activates both antigen-specific lymphoproliferation and CTLs to several antigens.

Experiments performed at Smith Kline Beecham Biologicals have demonstrated a clear synergistic effect of combinations of 3D-MPL and QS21 in the induction of both humoral and TH1 type cellular immune responses.

**The oil/water emulsion** is composed of 2 oils (a tocopherol and squalene), and of PBS containing Tween 80 as emulsifier. The emulsion comprised 5% squalene 5%

tocopherol 0.4% Tween 80 and had an average particle size of 180 nm (see WO 95/17210).

Experiments performed at Smith Kline Beecham Biologicals have proven that the  
5 adjunction of this O/W emulsion to 3D-MPL/QS21 further increases their immunostimulant properties.

#### **Preparation of the oil/water emulsion (2 fold concentrate)**

10 Tween 80 is dissolved in phosphate buffered saline (PBS) to give a 2% solution in the PBS. To provide 100ml two fold concentrate emulsion 5g of DL alpha tocopherol and 5ml of squalene are vortexed to mix thoroughly. 90ml of PBS/Tween solution is added and mixed thoroughly. The resulting emulsion is then passed through a syringe and finally microfluidised by using an M110S microfluidics machine. The resulting  
15 oil droplets have a size of approximately 180 nm.

#### **Preparation of oil in water formulation.**

Antigen prepared in accordance with example 1 or 2 (5µg) was diluted in 10 fold  
20 concentrated PBS pH 6.8 and H<sub>2</sub>O before consecutive addition of SB62, 3D-MPL (5µg), QS21 (5µg) and 50 µg/ml thiomersal as preservative at 5 min interval. The emulsion volume is equal to 50% of the total volume (50µl for a dose of 100µl).

All incubations were carried out at room temperature with agitation.  
25

### **6. IMMUNOGENICITY OF Tat AND Nef-Tat IN RODENTS**

Characterization of the immune response induced after immunization with Tat and  
30 NefTat was carried out. To obtain information on isotype profiles and cell-mediated immunity (CMI) two immunization experiments in mice were conducted. In the first experiment mice were immunized twice two weeks apart into the footpad with Tat or



- NefTat in the oxydized or reduced form, respectively. Antigens were formulated in an oil in water emulsion comprising squalene, tween 80<sup>™</sup> (polyoxyethylene sorbitan monooleate) QS21, 3D-MPL and  $\alpha$ -tocopherol, and a control group received the adjuvant alone. Two weeks after the last immunization sera were obtained and
- 5 subjected to Tat-specific ELISA (using reduced Tat for coating) for the determination of antibody titers and isotypes (Figure 6a). The antibody titers were highest in the mice having received oxydized Tat. In general, the oxydized molecules induced higher antibody titers than the reduced forms, and Tat alone induced higher antibody titers than NefTat. The latter observation was confirmed in the second experiment.
- 10 Most interestingly, the isotype profile of Tat-specific antibodies differed depending on the antigens used for immunization. Tat alone elicited a balanced IgG1 and IgG2a profile, while NefTat induced a much stronger T<sub>H2</sub> bias (Figure 6b). This was again confirmed in the second experiment.
- 15 In the second mouse experiment animals received only the reduced forms of the molecules or the adjuvant alone. Besides serological analysis (see above) lymphoproliferative responses from lymph node cells were evaluated. After restimulation of those cells in vitro with Tat or NefTat <sup>3</sup>H-thymidine incorporation was measured after 4 days of culture. Presentation of the results as stimulation indices
- 20 indicates that very strong responses were induced in both groups of mice having received antigen (Figure 7).

In conclusion, the mice studies indicate that Tat as well as Nef-Tat are highly immunogenic candidate vaccine antigens. The immune response directed against the

25 two molecules is characterized by high antibody responses with at least 50% IgG1. Furthermore, strong CMI responses (as measured by lymphoproliferation) were observed.

## 7. FUNCTIONAL PROPERTIES OF THE Tat AND Nef-Tat PROTEINS

30

The Tat and NefTat molecules in oxydized or reduced form were investigated for their ability to bind to human T cell lines. Furthermore, the effect on growth of

those cell lines was assessed. ELISA plates were coated overnight with different concentration of the Tat and NefTat proteins, the irrelevant gD from herpes simplex virus type II, or with a buffer control alone. After removal of the coating solution HUT-78 cells were added to the wells. After two hours of incubation the wells were  
5 washed and binding of cells to the bottom of the wells was assessed microscopically. As a quantitative measure cells were stained with toluidine blue, lysed by SDS, and the toluidine blue concentration in the supernatant was determined with an ELISA plate reader. The results indicate that all four proteins, Tat and NefTat in oxydized or reduced form mediated binding of the cells to the  
10 ELISA plate (Figure 8). The irrelevant protein (data not shown) and the buffer did not fix the cells. This indicates that the recombinantly expressed Tat-containing proteins bind specifically to human T cell lines.

In a second experiment HUT-78 cells were left in contact with the proteins for 16  
15 hours. At the end of the incubation period the cells were labeled with [ $^3\text{H}$ ]-thymidine and the incorporation rate was determined as a measure of cell growth. All four proteins included in this assay inhibited cell growth as judged by diminished radioactivity incorporation (Figure 9). The buffer control did not mediate this effect. These results demonstrate that the recombinant Tat-containing  
20 proteins are capable of inhibiting growth of a human T cell line.

In summary the functional characterization of the Tat and NefTat proteins reveals that these proteins are able to bind to human Tcell lines. Furthermore, the proteins are able to inhibit growth of such cell lines.

M 08.12.99

## CLAIMS

1. A vaccine composition which comprises a protein comprising
  - (a) an HIV Tat protein or derivative thereof linked to either (i) a fusion partner  
5 or (ii) an HIV Nef protein or derivative thereof; or
  - (b) an HIV Nef protein or derivative thereof linked to either (i) a fusion partner  
or (ii) an HIV Tat protein or derivative thereof; or
  - (c) an HIV Nef protein or derivative thereof linked to an HIV Tat protein or  
derivative thereof and a fusion partner,
- 10 in admixture with a pharmaceutically acceptable excipient.
  
2. A composition as claimed in claim 1 comprising a Tat-Nef fusion protein or  
derivative thereof.
  
- 15 3. A composition as claimed in claim 1 comprising a Nef-Tat fusion protein or  
derivative thereof.
  
4. A composition according to any one of claims 1 to 3 wherein the derivative  
of the Tat protein is a mutated Tat protein.
- 20 5. A composition according to any one of claims 1 to 4 wherein the derivative  
of the Nef protein is a mutated Nef protein.
  
6. A composition as claimed in any one of claims 1 - 5 wherein the fusion  
25 partner is a lipoprotein or derivative thereof.
  
7. A composition as claimed in claim 6 wherein the lipoprotein is Haemophilus  
Influenza B protein D or derivative thereof.
  
- 30 8. A composition as claimed in claim 7 wherein the fusion partner comprises  
between 100-130 amino acid from the N terminal of Haemophilus Influenza  
B protein D.

M 08.12.99

9. A composition as claimed in any one of Claims 1 to 8, wherein the Tat protein is the entire Tat protein.
- 5 10. A composition as claimed in any one of Claims 1 to 8, wherein the Nef protein is the entire Nef protein.
11. A composition as claimed in any one of Claims 1 to 10, wherein the Tat protein is fused to an HIV Nef protein and a fusion partner.
- 10 12. A composition as claimed in any one of claims 1 to 11, wherein the protein has a Histidine tail.
13. A composition as claimed in any one of claims 1 to 12 wherein the protein is carboxymethylated.
- 15 14. A composition as claimed in any one of claims 1 to 13, additionally comprising an adjuvant.
- 20 15. A composition as claimed in claim 14, wherein the adjuvant is a TH1 inducing adjuvant.
16. A composition as claimed in claim 14 or 15 which adjuvant comprises monophosphoryl lipid A or a derivative thereof such as 3 de-O-acylated monophosphoryl lipid A.
- 25 17. A composition as claimed in any one of claims 14 to 16 additionally comprising a saponin adjuvant.
- 30 18. A composition as claimed in any one of claims 14 to 17 which additionally comprises an oil in water emulsion.

NO. 10.00

19. A composition as claimed in any one of claims 1 to 18 further comprising HIV gp160 or its derivative gp120.
20. A protein comprising an HIV Tat protein or derivative thereof linked to an HIV Nef protein or derivative thereof in Nef-Tat or Tat-Nef orientation.
21. A nucleic acid encoding a protein of claim 20.
22. A host transformed with a nucleic acid of claim 21.
23. A host as claimed in claim 22 wherein the host is either *E.coli* or *Pichia pastoris*.
24. A method of producing a protein of claim 20, comprising providing a host as claimed in claim 22 or 23, expressing said protein and recovering the protein.
25. A method of preparing (i) an HIV Nef protein or derivative thereof or (ii) an HIV Tat protein or derivative thereof in *Pichia pastoris* which method comprises the steps of transforming *Pichia pastoris* with DNA encoding said HIV Nef protein or derivative thereof or HIV Tat protein or derivative thereof, expressing said protein and recovering the protein.
26. The method of claim 24 or claim 25 further comprising a carboxymethylation step performed on the expressed protein.
27. A method of producing a vaccine, comprising admixing the protein from any one of claims 24 to 26 with a pharmaceutically acceptable diluent.
28. The method of claim 27 further comprising the addition of HIV gp160 or its derivative gp120.

AMENDED SHEET

M 08.10.00

29. The method of claims 24 to 28 further comprising the addition of an adjuvant, particularly a TH1 inducing adjuvant.
30. A vaccine composition comprising a recombinant Tat-containing protein formulated with a mixture of 3D-MPL, QS21 and an oil in water emulsion
31. A composition as claimed in claim 30 wherein the oil in water emulsion comprises squalene, polyoxyethylene sorbitan monooleate and  $\alpha$ -tocopherol.

10

15

AMENDED SHEET

## (1) GENERAL INFORMATION

- (i) APPLICANT: SmithKline Beecham Biologicals S.A.
- (ii) TITLE OF THE INVENTION: Vaccine
- (iii) NUMBER OF SEQUENCES: 27
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: SmithKline Beecham
  - (B) STREET: Two New Horizons Court
  - (C) CITY: Brentford
  - (D) STATE:
  - (E) COUNTRY: Middx, UK
  - (F) ZIP: TW8 9EP
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
  - (B) FILING DATE: 26-SEP-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER:
  - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Bor, Fiona R
  - (B) REGISTRATION NUMBER:
  - (C) REFERENCE/DOCKET NUMBER:
- (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: 0181 975 2817
  - (B) TELEFAX: 0181 975 6141
  - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- SUBSTITUTE SHEET (RULE 26)**

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATCGTCCATG .GGT.GGC.A AG.TGG.T

28

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGGCTACTAG TGCAGTTCTT GAA

23

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCGTACTAG T.GAG.CCA. GTA.GAT.C

29

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGCTACTAG TTTCCTTCGG GCCT

24

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATCGTCCATG GAGCCAGTAG ATC

23



(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 441 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

ATGGATCCAA	AAACTTTAGC	CCTTTCCTTA	TTAGCAGCTG	GCGTACTAGC	AGGTTGTAGC	60
AGCCATT CAT	CAAATATGGC	GAATACCCAA	ATGAAATCAG	ACAAAATCAT	TATTGCTCAC	120
CGTGGTGCTA	GCGGTTATTT	ACCAGAGCAT	ACGTTAGAAT	CTAAAGCACT	TGCTTTTGCA	180
CAACAGGCTG	ATTATTTAGA	GCAAGATTTA	GCAATGACTA	AGGATGGTCG	TTTAGTGGTT	240
ATTACAGATC	ACTTTTTTAGA	TGGCTTGACT	GATGTTGCGA	AAAAATTCCC	ACATCGTCAT	300
CGTAAAGATG	GCCGTTACTA	TGTCATCGAC	TTTACCTTAA	AAGAAATTCA	AAGTTTAGAA	360
ATGCAGAA	ACTTTGAAAC	CATGGCCACG	TGTGATCAGA	GCTCAACTAG	TGGCCACCAT	420
CACCATCACC	ATTAATCTAG	A				441

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

Met	Asp	Pro	Lys	Thr	Leu	Ala	Leu	Ser	Leu	Leu	Ala	Ala	Gly	Val	Leu
1				5					10					15	
Ala	Gly	Cys	Ser	Ser	His	Ser	Ser	Asn	Met	Ala	Asn	Thr	Gln	Met	Lys
			20					25					30		
Ser	Asp	Lys	Ile	Ile	Ile	Ala	His	Arg	Gly	Ala	Ser	Gly	Tyr	Leu	Pro
		35					40					45			
Glu	His	Thr	Leu	Glu	Ser	Lys	Ala	Leu	Ala	Phe	Ala	Gln	Gln	Ala	Asp
	50					55				60					
Tyr	Leu	Glu	Gln	Asp	Leu	Ala	Met	Thr	Lys	Asp	Gly	Arg	Leu	Val	Val
65					70					75					80
Ile	His	Asp	His	Phe	Leu	Asp	Gly	Leu	Thr	Asp	Val	Ala	Lys	Lys	Phe
			85						90					95	
Pro	His	Arg	His	Arg	Lys	Asp	Gly	Arg	Tyr	Tyr	Val	Ile	Asp	Phe	Thr
			100					105					110		
Leu	Lys	Glu	Ile	Gln	Ser	Leu	Glu	Met	Thr	Glu	Asn	Phe	Glu	Thr	Met
		115					120					125			
Ala	Thr	Cys	Asp	Gln	Ser	Ser	Thr	Ser	Gly	His	His	His	His	His	His
	130					135					140				

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 648 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

ATGGGTGGCA	AGTGGTCAAA	AAGTAGTGTG	GTTGGATGGC	CTACTGTAAG	GGAAAAGAAATG	60
AGACGAGCTG	AGCCAGCAGC	AGATGGGGTG	GGAGCAGCAT	CTCGAGACCT	GGAAAAACAT	120
GGAGCAATCA	CAAGTAGCAA	TACAGCAGCT	ACCAATGCTG	CTTGTGCCTG	GCTAGAAGCA	180
CAAGAGGAGG	AGGAGGTGGG	TTTTCCAGTC	ACACCTCAGG	TACCTTTAAG	ACCAATGACT	240
TACAAGGCAG	CTGTAGATCT	TAGCCACTTT	TTAAAAGAAA	AGGGGGGACT	GGAAGGGCTA	300
ATTCACTCCC	AACGAAGACA	AGATATCCTT	GATCTGTGGA	TCTACCACAC	ACAAGGCTAC	360
TTCCCTGATT	GGCAGAACTA	CACACCAGGG	CCAGGGGTGA	GATATCCACT	GACCTTTGGA	420
TGGTGCTACA	AGCTAGTACC	AGTTGAGCCA	GATAAGGTAG	AAGAGGCCAA	TAAAGGAGAG	480
AACACGAGCT	TGTTACACCC	TGTGAGCCTG	CATGGAATGG	ATGACCCTGA	GAGAGAAGTG	540
TTAGAGTGGA	GGTTTGACAG	CCGCCTAGCA	TTTCATACAG	TGGCCCGAGA	GCTGCATCCG	600
GAGTACTTCA	AGAAGTGCAC	TAGTGGCCAC	CATCACCATC	ACCATTAA		648

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 216 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

Met	Gly	Gly	Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val
1				5					10					15	
Arg	Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala
			20					25					30		
Ala	Ser	Arg	Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr
		35					40					45			
Ala	Ala	Thr	Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	Glu
	50					55					60				
Glu	Val	Gly	Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr
65					70					75					80
Tyr	Lys	Ala	Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	Gly
				85					90					95	
Leu	Glu	Gly	Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu
			100					105					110		
Trp	Ile	Tyr	His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr
		115					120					125			
Pro	Gly	Pro	Gly	Val	Arg	Tyr	Pro	Leu	Thr	Phe	Gly	Trp	Cys	Tyr	Lys
	130					135					140				
Leu	Val	Pro	Val	Glu	Pro	Asp	Lys	Val	Glu	Glu	Ala	Asn	Lys	Gly	Glu
145					150					155					160
Asn	Thr	Ser	Leu	Leu	His	Pro	Val	Ser	Leu	His	Gly	Met	Asp	Asp	Pro
			165						170					175	
Glu	Arg	Glu	Val	Leu	Glu	Trp	Arg	Phe	Asp	Ser	Arg	Leu	Ala	Phe	His
			180					185					190		
His	Val	Ala	Arg	Glu	Leu	His	Pro	Glu	Tyr	Phe	Lys	Asn	Cys	Thr	Ser
		195					200						205		
Gly	His	His	His	His	His	His									
	210					215									

**SUBSTITUTE SHEET (RULE 26)**

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGAGCCAG	TAGATCCTAG	ACTAGAGCCC	TGGAAGCATC	CAGGAAGTCA	GCCTAAACT	60
GCTTGTTACCA	ATTGCTATTG	TAAAAGTGT	TGCTTTCATT	GCCAAGTTTG	TTTCATAACA	120
AAAGCCTTAG	GCATCTCCTA	TGGCAGGAAG	AAGCGGAGAC	AGCGACGAAG	ACCTCCTCAA	180
GGCAGTCAGA	CTCATCAAGT	TTCTCTATCA	AAGCAACCCA	CCTCCCAATC	CCGAGGGGAC	240
CCGACAGGCC	CGAAGGAAAC	TAGTGGCCAC	CATCACCATC	ACCATTAA		288

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met	Glu	Pro	Val	Asp	Pro	Arg	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	Ser
1				5					10					15	
Gln	Pro	Lys	Thr	Ala	Cys	Thr	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	Cys	Phe
			20					25					30		
His	Cys	Gln	Val	Cys	Phe	Ile	Thr	Lys	Ala	Leu	Gly	Ile	Ser	Tyr	Gly
		35					40					45			
Arg	Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Pro	Pro	Gln	Gly	Ser	Gln	Thr
		50				55					60				
His	Gln	Val	Ser	Leu	Ser	Lys	Gln	Pro	Thr	Ser	Gln	Ser	Arg	Gly	Asp
65					70					75				80	
Pro	Thr	Gly	Pro	Lys	Glu	Thr	Ser	Gly	His	His	His	His	His	His	
				85					90					95	

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 909 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGGTGGCA	AGTGGTCAAA	AAGTAGTGTG	GTTGGATGGC	CTACTGTAAG	GGAAAGAATG	60
AGACGAGCTG	AGCCAGCAGC	AGATGGGGTG	GGAGCAGCAT	CTCGAGACCT	GGAAAAACAT	120
GGAGCAATCA	CAAGTAGCAA	TACAGCAGCT	ACCAATGCTG	CTTGTGCCTG	GCTAGAAGCA	180
CAAGAGGAGG	AGGAGGTGGG	TTTCCAGTC	ACACCTCAGG	TACCTTTAAG	ACCAATGACT	240
TACAAGGCAG	CTGTAGATCT	TAGCCACTTT	TTAAAAGAAA	AGGGGGGACT	GGAAGGGCTA	300
ATTCATCCC	AACGAAGACA	AGATATCCTT	GATCTGTGGA	TCTACCACAC	ACAAGGCTAC	360

```

TTCCTGATT GGCAGAACTA CACACCAGGG CCAGGGGTCA GATATCCACT GACCTTTGGA 420
TGGTGCTACA AGCTAGTACC AGTTGAGCCA GATAAGGTAG AAGAGGCCAA TAAAGGAGAG 480
AACACCAGCT TGTTACACCC TGTGAGCCTG CATGGAATGG ATGACCCTGA GAGAGAAGTG 540
TTAGAGTGGA GGTTTGACAG CCGCCTAGCA TTTCATCACG TGGCCCGAGA GCTGCATCCG 600
GAGTACTTCA AGAACTGCAC TAGTGAGCCA GTAGATCCTA GACTAGAGCC CTGGAAGCAT 660
CCAGGAAGTC AGCCTAAAAC TGCTTGTACC AATTGCTATT GTAAAAAGTG TTGCTTTCAT 720
TGCCAAGTTT GTTTCATAAC AAAAGCCTTA GGCATCTCCT ATGGCAGGAA GAAGCGGAGA 780
CAGCGACGAA GACCTCCTCA AGGCAGTCAG ACTCATCAAG TTTCTCTATC AAAGCAACCC 840
ACCTCCCAAT CCCGAGGGGA CCCGACAGGC CCGAAGGAAA CTAGTGGCCA CCATCACCAT 900
CACCATTAA 909

```

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 303 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

Met Gly Gly Lys Trp Ser Lys Ser Ser Val Val Gly Trp Pro Thr Val
1          5          10          15
Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala
20          25          30
Ala Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr
35          40          45
Ala Ala Thr Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu
50          55          60
Glu Val Gly Phe Pro Val Thr Pro Gln Val Pro Leu Arg Pro Met Thr
65          70          75          80
Tyr Lys Ala Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly
85          90          95
Leu Glu Gly Leu Ile His Ser Gln Arg Arg Gln Asp Ile Leu Asp Leu
100          105          110
Trp Ile Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr
115          120          125
Pro Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Tyr Lys
130          135          140
Leu Val Pro Val Glu Pro Asp Lys Val Glu Glu Ala Asn Lys Gly Glu
145          150          155          160
Asn Thr Ser Leu Leu His Pro Val Ser Leu His Gly Met Asp Asp Pro
165          170          175
Glu Arg Glu Val Leu Glu Trp Arg Phe Asp Ser Arg Leu Ala Phe His
180          185          190
His Val Ala Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser
195          200          205
Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser Gln
210          215          220
Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe His
225          230          235          240
Cys Gln Val Cys Phe Ile Thr Lys Ala Leu Gly Ile Ser Tyr Gly Arg
245          250          255
Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Gly Ser Gln Thr His
260          265          270
Gln Val Ser Leu Ser Lys Gln Pro Thr Ser Gln Ser Arg Gly Asp Pro

```

7 / 15

275 280 285  
 Thr Gly Pro Lys Glu Thr Ser Gly His His His His His His  
 290 295 300

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1029 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATGGATCCAA AACTTTTAGC CCTTTCTTTA TTAGCAGCTG GCGTACTAGC AGGTTGTAGC 60  
 AGCCATTCAT CAAATATGGC GAATACCCAA ATGAAATCAG ACAAATCAT TATTGCTCAC 120  
 CGTGGTGCTA GCGGTTATTT ACCAGAGCAT ACGTTAGAAT CTAAAGCACT TGCTTTTGCA 180  
 CAACAGGCTG ATTATTTAGA GCAAGATTTA GCAATGACTA AGGATGGTCG TTTAGTGGTT 240  
 ATTCACGATC ACTTTTTTGA TGGCTTGA CTGATGTTGCGA AAAAATTCCC ACATCGTCAT 300  
 CGTAAAGATG GCCGTTACTA TGTCATCGAC TTTACCTTAA AAGAAATTCA AAGTTTAGAA 360  
 ATGACAGAAA ACTTTGAAAC CATGGGTGGC AAGTGGTCAA AAAGTAGTGT GGTGGATGG 420  
 CCTACTGTAA GGGAAAGAAT GAGACGAGCT GAGCCAGCAG CAGATGGGGT GGGAGCAGCA 480  
 TCTCGAGACC TGGAAAAACA TGGAGCAATC ACAAGTAGCA ATACAGCAGC TACCAATGCT 540  
 GCTTGTGCCT GGCTAGAAGC ACAAGAGGAG GAGGAGGTGG GTTTCCAGT CACACCTCAG 600  
 GTACCTTTAA GACCAATGAC TTACAAGGCA GCTGTAGATC TTAGCCACTT TTTAAAGAA 660  
 AAGGGGGGAC TGGGAAGGGCT AATTCCTCC CAACGAAGAC AAGATATCCT TGATCTGTGG 720  
 ATCTACCACA CACAAGGCTA CTTCCCTGAT TGGCAGA ACT ACACACCAGG GCCAGGGGTC 780  
 AGATATCCAC TGACCTTTGG ATGGTGCTAC AAGCTAGTAC CAGTTGAGCC AGATAAGGTA 840  
 GAAGAGGCCA ATAAAGGAGA GAACACCAGC TTGTTACACC CTGTGAGCCT GCATGGAATG 900  
 GATGACCCTG AGAGAGAAGT GTTAGAGTGG AGGTTTGACA GCCGCCTAGC ATTTTCATCAC 960  
 GTGGCCCGAG AGCTGCATCC GGAGTACTTC AAGAACTGCA CTAGTGGCCA CCATCACCAT 1020  
 CACCATTAA 1029

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 325 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Cys Ser Ser His Ser Ser Asn Met Ala Asn Thr Gln Met Lys Ser Asp  
 1 5 10 15  
 Lys Ile Ile Ile Ala His Arg Gly Ala Ser Gly Tyr Leu Pro Glu His  
 20 25 30  
 Thr Leu Glu Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala Asp Tyr Leu  
 35 40 45  
 Glu Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val Ile His  
 50 55 60  
 Asp His Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe Pro His  
 65 70 75 80  
 Arg His Arg Lys Asp Gly Arg Tyr Tyr Val Ile Asp Phe Thr Leu Lys  
 85 90 95

Glu Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Glu Thr Met Gly Gly  
 100 105 110  
 Lys Trp Ser Lys Ser Ser Val Val Gly Trp Pro Thr Val Arg Glu Arg  
 115 120 125  
 Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala Ala Ser Arg  
 130 135 140  
 Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala Ala Thr  
 145 150 155 160  
 Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu Val Gly  
 165 170 175  
 Phe Pro Val Thr Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys Ala  
 180 185 190  
 Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Glu Gly  
 195 200 205  
 Leu Ile His Ser Gln Arg Arg Gln Asp Ile Leu Asp Leu Trp Ile Tyr  
 210 215 220  
 His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly Pro  
 225 230 235 240  
 Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Tyr Lys Leu Val Pro  
 245 250 255  
 Val Glu Pro Asp Lys Val Glu Glu Ala Asn Lys Gly Glu Asn Thr Ser  
 260 265 270  
 Leu Leu His Pro Val Ser Leu His Gly Met Asp Asp Pro Glu Arg Glu  
 275 280 285  
 Val Leu Glu Trp Arg Phe Asp Ser Arg Leu Ala Phe His His Val Ala  
 290 295 300  
 Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser Gly His His  
 305 310 315 320  
 His His His His

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATGGATCCAA	AAACTTTAGC	CCTTTCTTTA	TTAGCAGCTG	GCGTACTAGC	AGGTTGTAGC	60
AGCCATTCAT	CAAATATGGC	GAATACCCAA	ATGAAATCAG	ACAAAATCAT	TATTGCTCAC	120
CGTGGTGCTA	GCGGTTATTT	ACCAGAGCAT	ACGTTAGAAT	CTAAAGCACT	TGCGTTTGCA	180
CAACAGGCTG	ATTATTTAGA	GCAAGATTTA	GCAATGACTA	AGGATGGTCG	TTTAGTGGTT	240
ATTCACGATC	ACTTTTTAGA	TGGCTTGACT	GATGTTGCGA	AAAAATTCCC	ACATCGTCAT	300
CGTAAAGATG	GCCGTTACTA	TGTCATCGAC	TTACCTTAA	AAGAAATTCA	AAGTTTAGAA	360
ATGACAGAAA	ACTTTGAAAC	CATGGGTGGC	AAGTGGTCAA	AAAGTAGTGT	GGTTGGATGG	420
CCTACTGTAA	GGGAAAGAAT	GAGACGAGCT	GAGCCAGCAG	CAGATGGGGT	GGGAGCAGCA	480
TCTCGAGACC	TGGAAAAACA	TGGAGCAATC	ACAAGTAGCA	ATACAGCAGC	TACCAATGCT	540
GCTTGTGCCT	GGCTAGAAGC	ACAAGAGGAG	GAGGAGGTGG	GTTTTCCAGT	CACACCTCAG	600
GTACCTTTAA	GACCAATGAC	TTACAAGGCA	GCTGTAGATC	TTAGCCACTT	TTTAAAAGAA	660
AAGGGGGGAC	TGGAAGGGCT	AATTCACCTC	CAACGAAGAC	AAGATATCCT	TGATCTGTGG	720
ATCTACCACA	CACAAGGCTA	CTTCCCTGAT	TGGCAGAACT	ACACACCAGG	GCCAGGGGTC	780
AGATATCCAC	TGACCTTTGG	ATGGTGCTAC	AAGCTAGTAC	CAGTTGAGCC	AGATAAGGTA	840
GAAGAGGCCA	ATAAAGGAGA	GAACACCAGC	TTGTTACACC	CTGTGAGCCT	GCATGGAATG	900

GATGACCCTG	AGAGAGAAGT	GTTAGAGTGG	AGGTTTGACA	GCCGCCTAGC	ATTTTCATCAC	960
GTGGCCCCGAG	AGCTGCATCC	GGAGTACTTC	AAGAACTGCA	CTAGTGAGCC	AGTAGATCCT	1020
AGACTAGAGC	CCTGGAAGCA	TCCAGGAAGT	CAGCCTAAAA	CTGCTTGAC	CAATTGCTAT	1080
TGTAAAAAGT	GTTGCTTTCA	TTGCCAAGTT	TGTTTCATAA	CAAAAGCCTT	AGGCATCTCC	1140
TATGGCAGGA	AGAAGCGGAG	ACAGCGACGA	AGACCTCCTC	AAGGCAGTCA	GACTCATCAA	1200
GTTTCTCTAT	CAAAGCAACC	CACCTCCCAA	TCCCGAGGGG	ACCCGACAGG	CCCGAAGGAA	1260
ACTAGTGGCC	ACCATCACCA	TCACCATTAA				1290

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Cys	Ser	Ser	His	Ser	Ser	Asn	Met	Ala	Asn	Thr	Gln	Met	Lys	Ser	Asp	1	5	10	15
Lys	Ile	Ile	Ile	Ala	His	Arg	Gly	Ala	Ser	Gly	Tyr	Leu	Pro	Glu	His	20	25	30	
Thr	Leu	Glu	Ser	Lys	Ala	Leu	Ala	Phe	Ala	Gln	Gln	Ala	Asp	Tyr	Leu	35	40	45	
Glu	Gln	Asp	Leu	Ala	Met	Thr	Lys	Asp	Gly	Arg	Leu	Val	Val	Ile	His	50	55	60	
Asp	His	Phe	Leu	Asp	Gly	Leu	Thr	Asp	Val	Ala	Lys	Lys	Phe	Pro	His	65	70	75	80
Arg	His	Arg	Lys	Asp	Gly	Arg	Tyr	Tyr	Val	Ile	Asp	Phe	Thr	Leu	Lys	85	90	95	
Glu	Ile	Gln	Ser	Leu	Glu	Met	Thr	Glu	Asn	Phe	Glu	Thr	Met	Gly	Gly	100	105	110	
Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val	Arg	Glu	Arg	115	120	125	
Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala	Ala	Ser	Arg	130	135	140	
Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr	Ala	Ala	Thr	145	150	155	160
Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	Glu	Glu	Val	Gly	165	170	175	
Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr	Tyr	Lys	Ala	180	185	190	
Ala	Val	Asp	Leu	Ser	His	Phe	Leu	Lys	Glu	Lys	Gly	Gly	Leu	Glu	Gly	195	200	205	
Leu	Ile	His	Ser	Gln	Arg	Arg	Gln	Asp	Ile	Leu	Asp	Leu	Trp	Ile	Tyr	210	215	220	
His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr	Pro	Gly	Pro	225	230	235	240
Gly	Val	Arg	Tyr	Pro	Leu	Thr	Phe	Gly	Trp	Cys	Tyr	Lys	Leu	Val	Pro	245	250	255	
Val	Glu	Pro	Asp	Lys	Val	Glu	Glu	Ala	Asn	Lys	Gly	Glu	Asn	Thr	Ser	260	265	270	
Leu	Leu	His	Pro	Val	Ser	Leu	His	Gly	Met	Asp	Asp	Pro	Glu	Arg	Glu	275	280	285	
Val	Leu	Glu	Trp	Arg	Phe	Asp	Ser	Arg	Leu	Ala	Phe	His	His	Val	Ala	290	295	300	

10 / 15

```

Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser Glu Pro Val
305                      310                      315                      320
Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser Gln Pro Lys Thr
                      325                      330                      335
Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe His Cys Gln Val
                      340                      345                      350
Cys Phe Ile Thr Lys Ala Leu Gly Ile Ser Tyr Gly Arg Lys Lys Arg
                      355                      360                      365
Arg Gln Arg Arg Arg Pro Pro Gln Gly Ser Gln Thr His Gln Val Ser
                      370                      375                      380
Leu Ser Lys Gln Pro Thr Ser Gln Ser Arg Gly Asp Pro Thr Gly Pro
385                      390                      395                      400
Lys Glu Thr Ser Gly His His His His His His
                      405                      410

```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 981 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

ATGGATCCAA GCAGCCATTC ATCAAATATG GCGAATACCC AAATGAAATC AGACAAAATC      60
ATTATTGCTC ACCGTGGTGC TAGCGGTTAT TTACCAGAGC ATACGTTAGA ATCTAAAGCA      120
CTTGCGTTTG CACAACAGGC TGATTATTTA GAGCAAGATT TAGCAATGAC TAAGGATGGT      180
CGTTTAGTGG TTATTCACGA TCACTTTTTA GATGGCTTGA CTGATGTTGC GAAAAAATTC      240
CCACATCGTC ATCGTAAAGA TGGCCGTTAC TATGTCATCG ACTTTACCTT AAAAGAAATT      300
CAAAGTTTAG AAATGACAGA AAACCTTGAA ACCATGGGTG GCAAGTGGTC AAAAAGTAGT      360
GTGGTTGGAT GGCCTACTGT AAGGGAAAGA ATGAGACGAG CTGAGCCAGC AGCAGATGGG      420
GTGGGAGCAG CATCTCGAGA CCTGGAAAAA CATGGAGCAA TCACAAGTAG CAATACAGCA      480
GCTACCAATG CTGCTTGTGC CTGGCTAGAA GCACAAGAGG AGGAGGAGGT GGGTTTTCCA      540
GTCACACCTC AGGTACCTTT AAGACCAATG ACTTACAAGG CAGCTGTAGA TCTTAGCCAC      600
TTTTTAAAG AAAAGGGGGG ACTGGAAGGG CTAATTCAC CCCAACGAAG ACAAGATATC      660
CTTGATCTGT GGATCTACCA CACACAAGGC TACTTCCCTG ATTGGCAGAA CTACACACCA      720
GGGCCAGGGG TCAGATATCC ACTGACCTTT GGATGGTGCT ACAAGCTAGT ACCAGTTGAG      780
CCAGATAAGG TAGAAGAGGC CAATAAAGGA GAGAACACCA GCTTGTTACA CCCTGTGAGC      840
CTGCATGGAA TGGATGACCC TGAGAGAGAA GTGTTAGAGT GGAGGTTTGA CAGCCGCCTA      900
GCATTTTCATC ACGTGGCCCG AGAGCTGCAT CCGGAGTACT TCAAGAACTG CACTAGTGGC      960
CACCATCACC ATCACCATTA A

```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 327 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```

Met Asp Pro Ser Ser His Ser Ser Asn Met Ala Asn Thr Gln Met Lys
1                      5                      10                      15

```



Ser Asp Lys Ile Ile Ile Ala His Arg Gly Ala Ser Gly Tyr Leu Pro  
 20 25 30  
 Glu His Thr Leu Glu Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala Asp  
 35 40 45  
 Tyr Leu Glu Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val  
 50 55 60  
 Ile His Asp His Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe  
 65 70 75 80  
 Pro His Arg His Arg Lys Asp Gly Arg Tyr Tyr Val Ile Asp Phe Thr  
 85 90 95  
 Leu Lys Glu Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Glu Thr Met  
 100 105 110  
 Gly Gly Lys Trp Ser Lys Ser Ser Val Val Gly Trp Pro Thr Val Arg  
 115 120 125  
 Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala Ala  
 130 135 140  
 Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr Ala  
 145 150 155 160  
 Ala Thr Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu  
 165 170 175  
 Val Gly Phe Pro Val Thr Pro Gln Val Pro Leu Arg Pro Met Thr Tyr  
 180 185 190  
 Lys Ala Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu  
 195 200 205  
 Glu Gly Leu Ile His Ser Gln Arg Arg Gln Asp Ile Leu Asp Leu Trp  
 210 215 220  
 Ile Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro  
 225 230 235 240  
 Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Tyr Lys Leu  
 245 250 255  
 Val Pro Val Glu Pro Asp Lys Val Glu Glu Ala Asn Lys Gly Glu Asn  
 260 265 270  
 Thr Ser Leu Leu His Pro Val Ser Leu His Gly Met Asp Asp Pro Glu  
 275 280 285  
 Arg Glu Val Leu Glu Trp Arg Phe Asp Ser Arg Leu Ala Phe His His  
 290 295 300  
 Val Ala Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser Gly  
 305 310 315 320  
 His His His His His  
 325

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1242 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGGATCCAA	GCAGCCATTC	ATCAAATATG	GCGAATACCC	AAATGAAATC	AGACAAAATC	60
ATTATTGCTC	ACCGTGGTGC	TAGCGGTTAT	TTACCAGAGC	ATACGTTAGA	ATCTAAAGCA	120
CTTGCGTTTG	CACAACAGGC	TGATTATTTA	GAGCAAGATT	TAGCAATGAC	TAAGGATGGT	180
CGTTTAGTGG	TTATTCACGA	TCACTTTTTA	GATGGCTTGA	CTGATGTTGC	GAAAAAATTC	240
CCACATCGTC	ATCGTAAAGA	TGGCCGTTAC	TATGTCATCG	ACTTTACCTT	AAAAGAAATT	300

(2) INFORMATION FOR SEO ID NO:21:

(A) LENGTH: 414 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

Met 1	Asp	Pro	Ser	Ser 5	His	Ser	Ser	Asn	Met 10	Ala	Asn	Thr	Gln	Met 15	Lys
Ser	Asp	Lys	Ile 20	Ile	Ile	Ala	His	Arg 25	Gly	Ala	Ser	Gly	Tyr 30	Leu	Pro
Glu	His	Thr 35	Leu	Glu	Ser	Lys	Ala 40	Leu	Ala	Phe	Ala	Gln 45	Gln	Ala	Asp
Tyr	Leu 50	Glu	Gln	Asp	Leu	Ala 55	Met	Thr	Lys	Asp	Gly 60	Arg	Leu	Val	Val
Ile 65	His	Asp	His	Phe 70	Leu	Asp	Gly	Leu	Thr	Asp 75	Val	Ala	Lys	Lys	Phe 80
Pro	His	Arg	His 85	Arg	Lys	Asp	Gly	Arg	Tyr 90	Tyr	Val	Ile	Asp	Phe 95	Thr
Leu	Lys	Glu	Ile 100	Gln	Ser	Leu	Glu	Met	Thr	Glu	Asn	Phe	Glu 110	Thr	Met
Gly	Gly	Lys 115	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro 125	Thr	Val	Arg
Glu	Arg 130	Met	Arg	Arg	Ala	Glu 135	Pro	Ala	Ala	Asp	Gly 140	Val	Gly	Ala	Ala
Ser 145	Arg	Asp	Leu	Glu 150	Lys	His	Gly	Ala	Ile	Thr 155	Ser	Ser	Asn	Thr	Ala 160
Ala	Thr	Asn	Ala 165	Ala	Cys	Ala	Trp	Leu	Glu 170	Ala	Gln	Glu	Glu 175	Glu	Glu
Val	Gly	Phe 180	Pro	Val	Thr	Pro	Gln 185	Val	Pro	Leu	Arg	Pro 190	Met	Thr	Tyr
Lys	Ala 195	Ala	Val	Asp	Leu	Ser	His 200	Phe	Leu	Lys	Glu	Lys 205	Gly	Gly	Leu
Glu	Gly 210	Leu	Ile	His	Ser	Gln 215	Arg	Arg	Gln	Asp	Ile 220	Leu	Asp	Leu	Trp
Ile	Tyr	His	Thr	Gln	Gly	Tyr	Phe	Pro	Asp	Trp	Gln	Asn	Tyr	Thr	Pro

(2) INFORMATION FOR SEO ID NO:22:

(A) LENGTH: 288 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

ATGGAGCCAG	TAGATCTAG	ACTAGAGCCC	TGGAAGCATC	CAGGAAGTCA	GCCTAAAACT	60
GCTTGTACCA	ATTGCTATTG	TAAAAAGTGT	TGCTTTCATT	GCCAAGTTTG	TTTCATAACA	120
GCTGCCTTAG	GCATCTCCTA	TGGCAGGAAG	AAGCGGAGAC	AGCGACGAAG	ACCTCCTCAA	180
GGCAGTCTAGA	CTCATCAAGT	TTCTCTATCA	AAGCAACCCA	CCTCCCCAATC	CAAAGGGGGAG	240
CCGACAGGCC	CGAAGGAAAC	TAGTGGCCAC	CATCACCATC	ACCATTAAC		288

(A) LENGTH: 96 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

Met 1	Glu	Pro	Val	Asp 5	Pro	Arg	Leu	Glu	Pro 10	Trp	Lys	His	Pro	Gly 15	Ser
Gln	Pro	Lys	Thr 20	Ala	Cys	Thr	Asn	Cys 25	Tyr	Cys	Lys	Lys	Cys 30	Cys	Phe
His	Cys	Gln	Val	Cys	Phe	Ile	Thr	Ala	Ala	Leu	Gly	Ile	Ser	Tyr	Gly

(2) INFORMATION FOR SEQ ID NO:24:

(A) LENGTH: 909 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

ATGGGTGGCA	AGTGGTCAAA	AAGTAGTGTG	GTTGGATGGC	CTACTGTAAG	GGAAAGAATG	60
AGACGAGCTG	AGCCAGCAGC	AGATGGGGTG	GGAGCAGCAT	CTCGAGACCT	GGAAAAACAT	120
GGAGCAATCA	CAAGTAGCAA	TACAGCAGCT	ACCAATGCTG	CTTGTGCCTG	GCTAGAAGCA	180
CAAGAGGAGG	AGGAGGTGGG	TTTTCCAGTC	ACACCTCAGG	TACCTTTAAG	ACCAATGACT	240
TACAAGGCAG	CTGTAGATCT	TAGCCACTTT	TTAAAGAGAA	AGGGGGGACT	GGAAGGGCTA	300
ATTCACTCCC	AACGAAGACA	AGATATCCTT	GATCTGTGGA	TCTACCACAC	ACAAGGCTAC	360
TTCCCTGATT	GGCAGAACTA	CACACCAGGG	CCAGGGGTCA	GATATCCACT	GACCTTTGGA	420
TGGTGCTACA	AGCTAGTACC	AGTTGAGCCA	GATAAGGTAG	AAGAGGCCAA	TAAAGGAGAG	480
AACACCAGCT	TGTTACACCC	TGTGAGCCTG	CATGGAATGG	ATGACCCTGA	GAGAGAAGTG	540
TTAGAGTGGA	GGTTTGACAG	CCGCCTAGCA	TTTCATCACG	TGGCCCGAGA	GCTGCATCCG	600
GAGTACTTCA	AGAACTGCAC	TAGTGAGCCA	GTAGATCCTA	GACTAGAGCC	CTGGAAGCAT	660
CCAGGAAGTC	AGCCTAA AAC	TGCTTGTAAC	AATTGCTATT	GTAAAAAGTG	TTGCTTTCAT	720
TGCCAAGTTT	GTTTCATAAC	AGCTGCCTTA	GGCATCTCCT	ATGGCAGGAA	GAAGCGGAGA	780
CAGCGACGAA	GACCTCCTCA	AGGCAGTCAG	ACTCATCAAG	TTTCTCTATC	AAAGCAACCC	840
ACCTCCCAAT	CCAAGGGGA	GCCGACAGGC	CCGAAGGAAA	CTAGTGGCCA	CCATCACCAT	900
CACCATTA						909

(2) INFORMATION FOR SEO ID NO:25:

(A) LENGTH: 303 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEO ID NO:25:

Met	Gly	Gly	Lys	Trp	Ser	Lys	Ser	Ser	Val	Val	Gly	Trp	Pro	Thr	Val
1				5					10					15	
Arg	Glu	Arg	Met	Arg	Arg	Ala	Glu	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala
			20					25					30		
Ala	Ser	Arg	Asp	Leu	Glu	Lys	His	Gly	Ala	Ile	Thr	Ser	Ser	Asn	Thr
		35				40						45			
Ala	Ala	Thr	Asn	Ala	Ala	Cys	Ala	Trp	Leu	Glu	Ala	Gln	Glu	Glu	Glu
	50					55					60				
Glu	Val	Gly	Phe	Pro	Val	Thr	Pro	Gln	Val	Pro	Leu	Arg	Pro	Met	Thr
65				70					75					80	

15 / 15

Tyr Lys Ala Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly  
                             85                            90                            95  
 Leu Glu Gly Leu Ile His Ser Gln Arg Arg Gln Asp Ile Leu Asp Leu  
                             100                            105                            110  
 Trp Ile Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr  
                             115                            120                            125  
 Pro Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Tyr Lys  
                             130                            135                            140  
 Leu Val Pro Val Glu Pro Asp Lys Val Glu Glu Ala Asn Lys Gly Glu  
                             145                            150                            155                            160  
 Asn Thr Ser Leu Leu His Pro Val Ser Leu His Gly Met Asp Asp Pro  
                             165                            170                            175  
 Glu Arg Glu Val Leu Glu Trp Arg Phe Asp Ser Arg Leu Ala Phe His  
                             180                            185                            190  
 His Val Ala Arg Glu Leu His Pro Glu Tyr Phe Lys Asn Cys Thr Ser  
                             195                            200                            205  
 Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser Gln  
                             210                            215                            220  
 Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe His  
                             225                            230                            235                            240  
 Cys Gln Val Cys Phe Ile Thr Ala Ala Leu Gly Ile Ser Tyr Gly Arg  
                             245                            250                            255  
 Lys Lys Arg Arg Gln Arg Arg Arg Pro Gln Gly Ser Gln Thr His  
                             260                            265                            270  
 Gln Val Ser Leu Ser Lys Gln Pro Thr Ser Gln Ser Lys Gly Glu Pro  
                             275                            280                            285  
 Thr Gly Pro Lys Glu Thr Ser Gly His His His His His His  
                             290                            295                            300

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TTCGAAACCA TGGCCGCGGA CTAGTGGCCA CCATCACCAT CACCATTAC GGAATTC

57

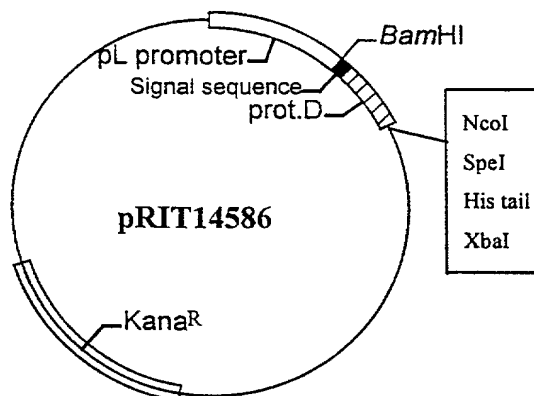
## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Thr Ser Gly His His His His His His  
 1                            5

**Figure 1:** A/ Map of plasmid pRIT14586

B/ Coding sequence of the first 127 amino acids  
of protein D and multiple cloning site. The signal  
sequence is underlined.

BamHI  
 ATG GAT CCA AAA ACT TTA GCC CTT TCT TTA TTA GCA GCT GGC GTA CTA GCA GGT TGT AGC AGC  
 Met Asp Pro Lys Thr Leu Ala Leu Ser Leu Leu Ala Ala Gly Val Leu Ala Gly Cys Ser Ser  
 CAT TCA TCA AAT ATG GCG AAT ACC CAA ATG AAA TCA GAC AAA ATC ATT ATT GCT CAC CGT GGT  
 His Ser Ser Asn Met Ala Asn Thr Gln Met Lys Ser Asp Lys Ile Ile Ile Ala His Arg Gly  
 GCT AGC GGT TAT TTA CCA GAG CAT ACG TTA GAA TCT AAA GCA CTT GCT TTT GCA CAA CAG GCT  
 Ala Ser Gly Tyr Leu Pro Glu His Thr Leu Glu Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala  
 GAT TAT TTA GAG CAA GAT TTA GCA ATG ACT AAG GAT GGT CGT TTA GTG GTT ATT CAC GAT CAC  
 Asp Tyr Leu Glu Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val Ile His Asp His  
 TTT TTA GAT GGC TTG ACT GAT GTT GCG AAA AAA TTC CCA CAT CGT CAT CGT AAA GAT GGC CGT  
 Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe Pro His Arg His Arg Lys Asp Gly Arg  
 TAC TAT GTC ATC GAC TTT ACC TTA AAA GAA ATT GAA AGT TTA GAA ATG ACA GAA AAC TTT GAA  
 Tyr Tyr Val Ile Asp Phe Thr Leu Lys Glu Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Glu  
NcoI SpeI XbaI  
 ACC ATG GCC ACG TGT GAT CAG AGC TCA ACT AGT GGA CAC CAT CAC CAT CAC CAT TAA TCT AGA  
 Thr Met Ala Thr Cys Asp Gln Ser Ser Thr Ser Gly His His His His His His \*

The amino acid sequence of Figure 1 relates to Seq. ID no. 7 and the nucleic acid sequence of  
Figure 1 relates to Seq. ID. No. 6.

The DNA and amino acid sequences of Nef-His; Tat-His; Nef-Tat-His fusion and mutated Tat is illustrated.

Pichia-expressed constructs (plain constructs)

⇒ Nef - HIS

DNA sequence (Seq. ID. No. 8)

ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGGAAAGA  
ATGAGACGAGCTGAGCCAGCAGCAGATGGGGTGGGAGCAGCATCTCGAGACCTGGAA  
AAACATGGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCCTGG  
CTAGAAGCACAAGAGGAGGAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTA  
AGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTTAAAAGAAAAGGGG  
GGACTGGAAGGGCTAATTCACCTCCCAACGAAGACAAGATATCCTTGATCTGTGGATC  
TACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTC  
AGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAG  
GTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCAT  
GGAATGGATGACCCTGAGAGAGAAGTGTAGAGTGGAGGTTTGACAGCCGCCTAGCA  
TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCACTAGTGGC  
CACCATCACCATCACCATTAA

Protein sequence (Seq. ID. No. 9)

MGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAITSSNTAATNAACAW  
LEAQEEEEVGFVTPQVPLRPMYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWI  
YHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPEPDKVEEANKGENTSLHPVSLH  
GMDDPEREVLEWRFD SRLAFHHVARELHPEYFKNCTSGHHHHHH.

⇒ Tat - HIS

DNA sequence (Seq. ID. No. 10)

ATGGAGCCAGTAGATCCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAA  
ACTGCTTGTACCAATTGCTATTGTAAAAAGTGTTGCTTTCATTGCCAAGTTTGTTC  
ATAACAAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGA  
CCTCCTCAAGGCAGTCAGACTCATCAAGTTTCTCTATCAAAGCAACCCACCTCCCAA

09509239.032300

TCCCGAGGGGACCCGACAGGCCCGAAGGAACTAGTGGCCACCATCACCATCACCAT  
TAA

Protein sequence (Seq. ID. No. 11)

MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYGRKKRRQRRR  
PPQGSQTHQVSLSKQPTSQSRGDPTGPKETSGHHHHHHH.

$\Rightarrow$  Nef - Tat - HIS

DNA sequence (Seq. ID. No. 12)

ATGGGTGGGCAAGTGGTGTTCAAAAAGTAGTGTGGTTGGATGGCCTACTGTGAAGGAAAGA  
ATGAGACGAGCTGAGCCAGCAGCAGATGGGGTGGGAGCAGCATCTCGAGACCTGGAA  
AAACATGGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCCTGG  
CTAGAAGCACAAAGAGGAGGAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTA  
AGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTTAAAAGAAAAGGGG  
GGACTGGAAGGGCTAATTCCTCCCAACGAAGACAAGATATCCTTGATCTGTGGATC  
TACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGT  
AGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAG  
GTAGAAGAGGCCAATAAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCAT  
GGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA  
TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCACTAGTGAG  
CCAGTAGATCCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAAACCTGCT  
TGTACCAATTGCTATTGTAAAAAGTGTTGCTTTTATTGCCAAGTTTGTTTCATAACA  
AAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCCT  
CAAGGCAGTCAGACTCATCAAGTTTCTCTATCAAAGCAACCCACCTCCCAATCCCGA  
GGGGACCCGACAGGCCCGAAGGAAACTAGTGGCCACCATCACCATCACCATTAA

*Protein sequence(Seq. ID. No. 13)*

MGGKWSKSSVVGWPTVRRERMRAEPAADGVGAASRDLEKHGAI TSSNTAATNAACAW  
 LEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDWI  
 YHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPEPDKVEEANKGENTSLLHPVSLH  
 GMDDPEREVLEWRFD SRLAFHHVARELHPEYFKNCTSEPVDPRLPEWKHPGSQPKTA  
 CTNCYCKKCCFHCQVCFITKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQPTSQSR  
 GDPTGPKETSGHHHHHH.

### E.coli-expressed constructs (fusion constructs)

$\Rightarrow$  *LipoD-Nef-HIS*





CSSHSSNMANTQMKSDKIIIAHRGASGYLPEHTLESKALAFQAQQADYLEQDLAMTKD  
GRLVVIHDHFLDGLTDVAKKFPHRHRKDGRIYVIDFTLKEIQSLEMTENFETMGGKW  
SKSSVVGWPTVRERMRAEPAADGVGAASRDLEKHGAITSSNTAATNAACAWLEAQE  
EEEVGFVPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDWIIYHTQG  
YFPDWQNYTPGPGVRYPLTFGWICYKLVPEPDKVEEANKGENTSLHHPVSLHGMDDP  
EREVLEWRFDSSLAFHHVARELHPEYFKNCTSEPVDPRLEPWKHPGSQPKTACTNCY  
CKKCCFHCQVCFITKALGISYGRKKRRQRRRPPQGSQTHQVLSLKQPTSQSRGDPTG  
PKETSGHHHHHHH.



7/17

Nucleotides corresponding to the Prot D Fusion Partner are in bold.

ATGGATCCAAGCAGCCATTCATCAAATATGGCGAATACCCAAATGAAATCAGACAAA  
 ATCATTATTGCTCACC GTGGT GCTAGCGGTTATTTACCAGAGCATACGTTAGAATCT  
 AAAGCACTTGC GTTTGCACAACAGGCTGATTATTTAGAGCAAGATTTAGCAATGACT  
 AAGGATGGTCGTTTAGTGGTTATTACGATCACTTTTAGATGGCTTGACTGATGTT  
 GCGAAAAAATCCACATCGTCATCGTAAAGATGGCCGTACTATGTCATCGACTTT  
 ACCTTAAAAGAAATTCAAAGTTTAGAAATGACAGAAAACCTTTGAAACCATGGGTGGC  
 AAGTGGTCAAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGGAAAGAATGAGACGA  
 GCTGAGCCAGCAGCAGATGGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGA  
 GCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCCTGGCTAGAAGCA  
 CAAGAGGAGGAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATG  
 ACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAA  
 GGGCTAATTCCTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACA  
 CAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTGAGATATCCA  
 CTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAG  
 GCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGGAT  
 GACCCTGAGAGAGAAGTGTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTTCATCAC  
 GTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCACTAGTGAGCCAGTAGAT  
 CCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAACTGCTTGTACCAAT  
 TGCTATTGTAAAAGTGTTGCTTTCATTGCCAAGTTTGTTCATAACAAAAGCCTTA  
 GGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCCTCAAGGCAGT  
 CAGACTCATCAAGTTTCTCTATCAAAGCAACCCACCTCCCAATCCCGAGGGGACCCG  
 ACAGGCCCCGAAGGAAACTAGTGGCCACCATCACCATCACCATTAA

Protein sequence (Seq. ID. No. 21)

(Amino-acids corresponding to Prot D fusion partner are in bold)

MDPSSHSSNMANTQMKSDKIIIAHRGASGYLPEHTLESKALAFQAQADYLEQDLAMT  
 KDGRLLVVIHDHFLDGLTDVAKKFPHRHRKDGRYYVIDFTLKEIQSLEMTENFETMGG  
 KWSKSSVVGWPTVRERMRRRAEPAADGVGAASRDLEKHGAITSSNTAATNAACAWLEA  
 QEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDWLYHT  
 QGYFPDWQNYTPGPGVRYPLTFGWICYKLVPEPDKVEEANKGENTSLHHPVSLHGMD  
 DPEREVLEWRFD SRLAFHHVARELHPEYFKNCTSEPVDPRLEPWKHPGSQPKTACTN  
 CYCKKCCFHCQVCFITKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQPTSQSRGDP  
 TGPKETSGHHHHHH.

⇒ Tat-MUTANT-HIS

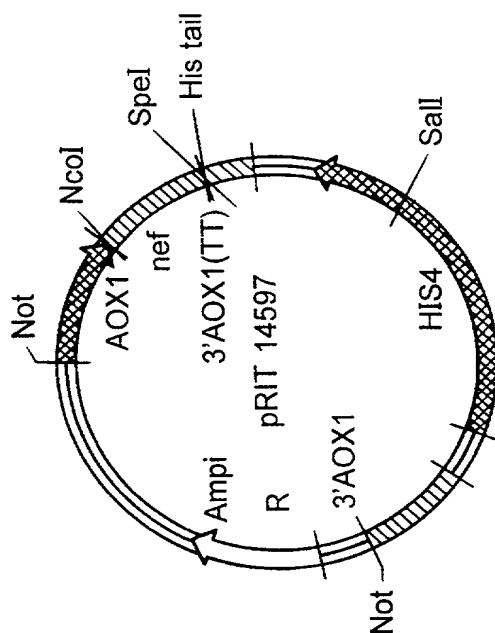
DNA sequence (Seq. ID. No. 22)



Mutated amino-acids in Tat sequence are in bold.

MGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKH	40
GAITSSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMT	80
YKAAVDLSHFLKEKGGLLEGLIHSQRRQDILDLWIYHTQGY	120
FPDWQNYTPGPGVRYPLTFGWICYKLVPEPDKVEEANKGE	160
NTSL LHVPVSLHGMDPPEREVLEWRFD SRLAFHHVARELHP	200
EYFKNCTSEPVDPRLPEPWKHPGSQPKTACTNCYCKKCCFH	240
CQVCFITAAALGISYGRKKRRQRRRPPQGSQTHQVSLSKQP	280
TSQSKGEPTGPKETSGHHHHHHH.	302

[illegible]

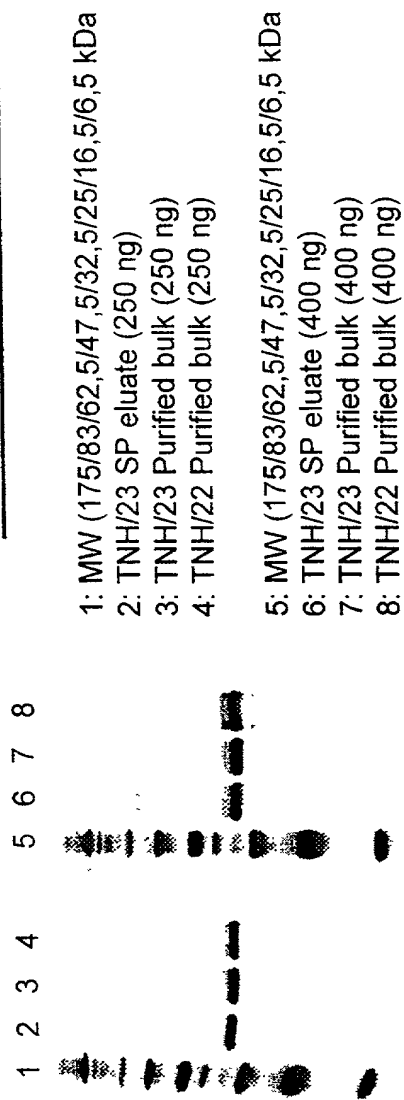
[illegible]

MCS POLYLINKER: *nef* gene inserted between NcoI and SpeI sites.

<i>Acu II</i>	<i>Nco I</i>	<i>Spe I</i>	<i>Eco RI</i>
TTCGAA	<u>ACC.ATGCCGCGGACTAGT</u>	GGC.CAC.CAT.CAC.CAT.CAC.CAT.CAC.CAT.TAA.CGGAATTC	
		Thr . Ser . Gly . His . His . His . His . His . His	

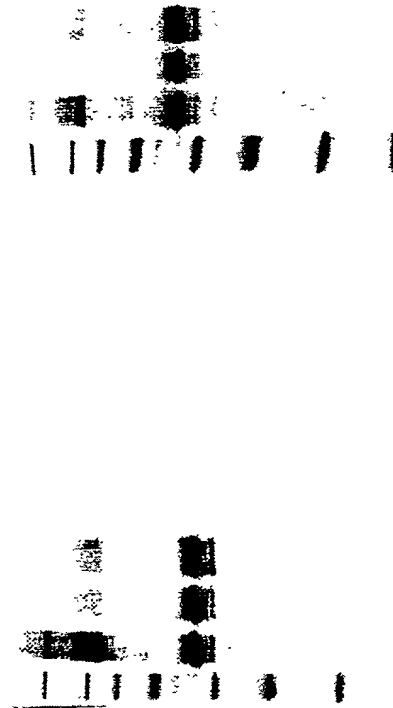
The amino acid sequence of Figure 3 relates to Seq. ID no. 27 and the nucleic acid sequence of Figure 3 relates to Seq. ID. No.26.

**Fig . 4** SDS-PAGE: Nef-Tat-his fusion protein



Daiichi Silver Staining

1 2 3 4

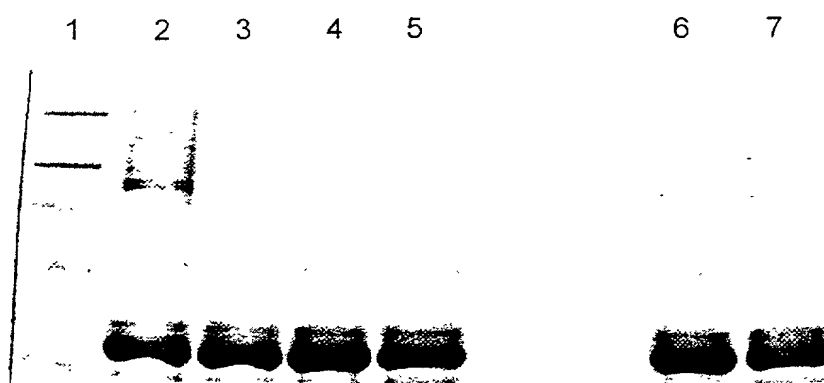


Blot Tat2

Blot $\alpha$ Nef-Tat (LAS 97340)



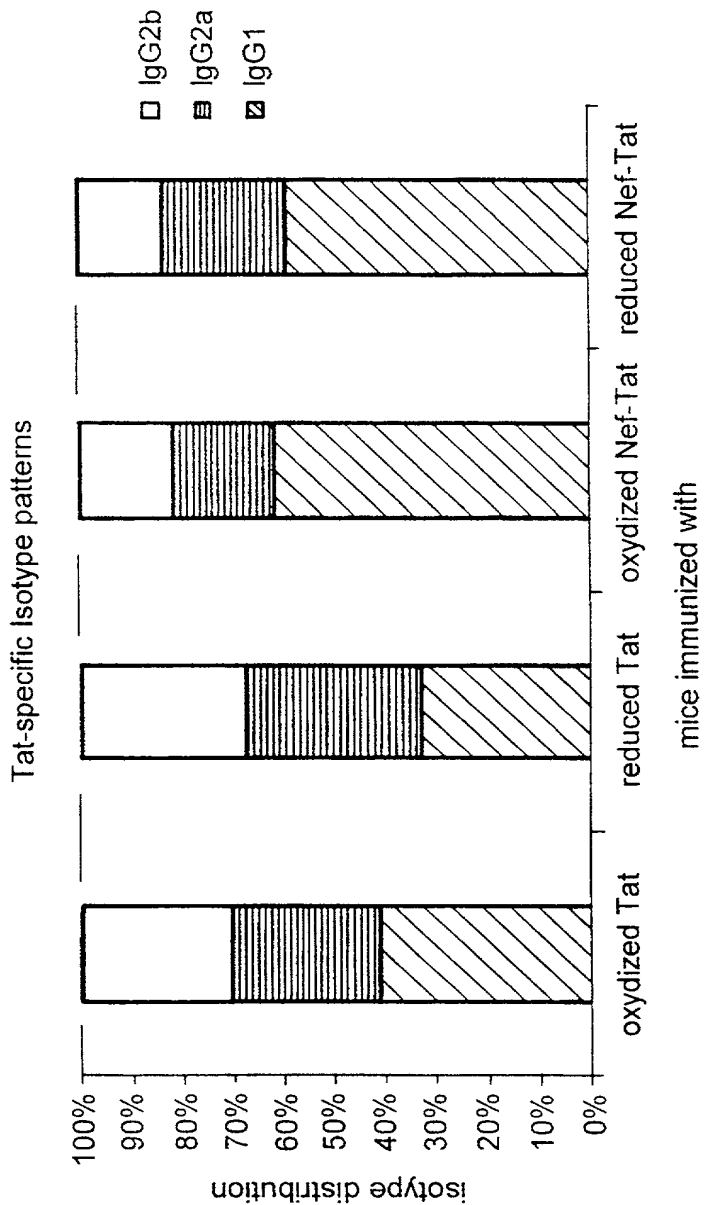
12/17

**Fig . 5** SDS-PAGE: Nef-Tat-his fusion proteinCoomassie blue G250

- 1: MW (175/83/62,5/47,5/32,5/25/16,5/6,5 kDa)
- 2: TNH/23 SP eluate (4 µg)
- 3: TNH/23 Superdex200 eluate (4 µg)
- 4: TNH/23 Purified bulk (4 µg)
- 5: TNH/22 Purified bulk (4 µg)
- 6: TNH/23 Purified bulk (4 µg) / non reducing conditions
- 7: TNH/22 Purified bulk (4 µg) / non reducing conditions

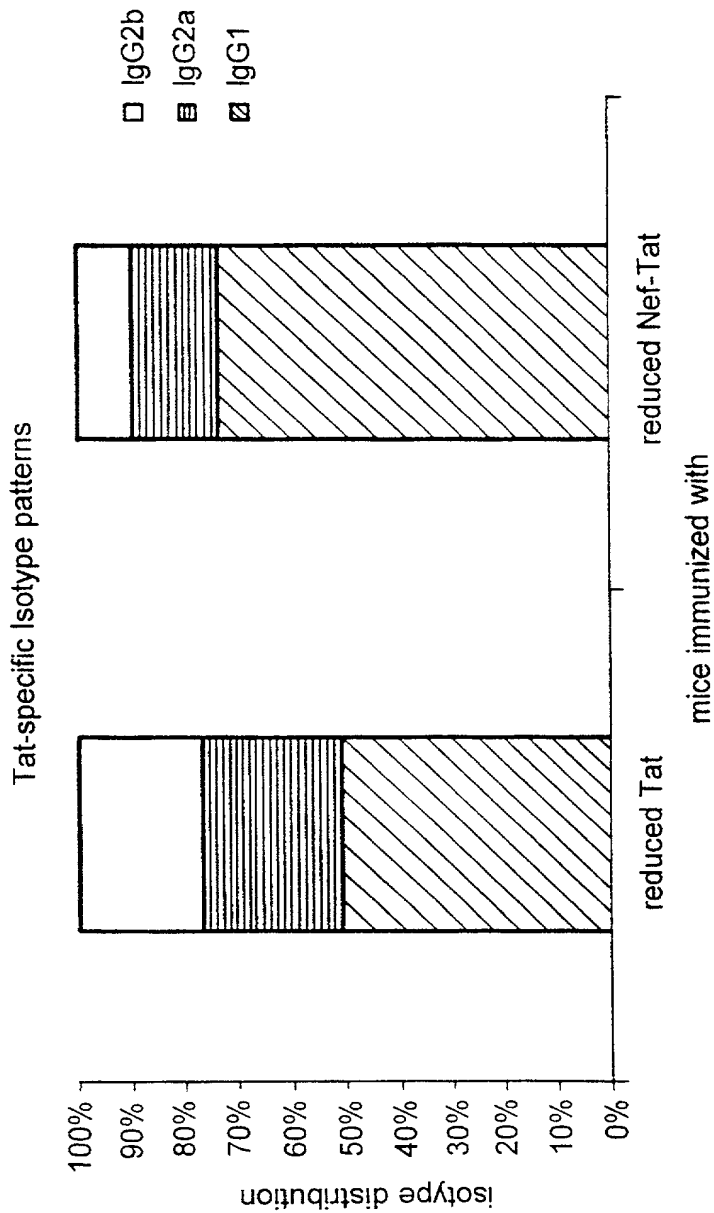
**Fig. 6A** Tat-specific antibody titers and isotypes

group	immunization	midpoint titers					ratio IgG1/IgG2a
		Ig	IgG1	IgG2a	IgG2b		
1	oxydized Tat	353557	135538	98771	98763	1,372	
2	reduced Tat	252275	72087	76273	72014	0,945	
3	oxydized Nef-Tat	246466	179616	60835	53563	2,953	
4	reduced Nef-Tat	91726	73767	30948	20679	2,384	
5	adjuvant only	<4000	<4000	<4000	<4000		



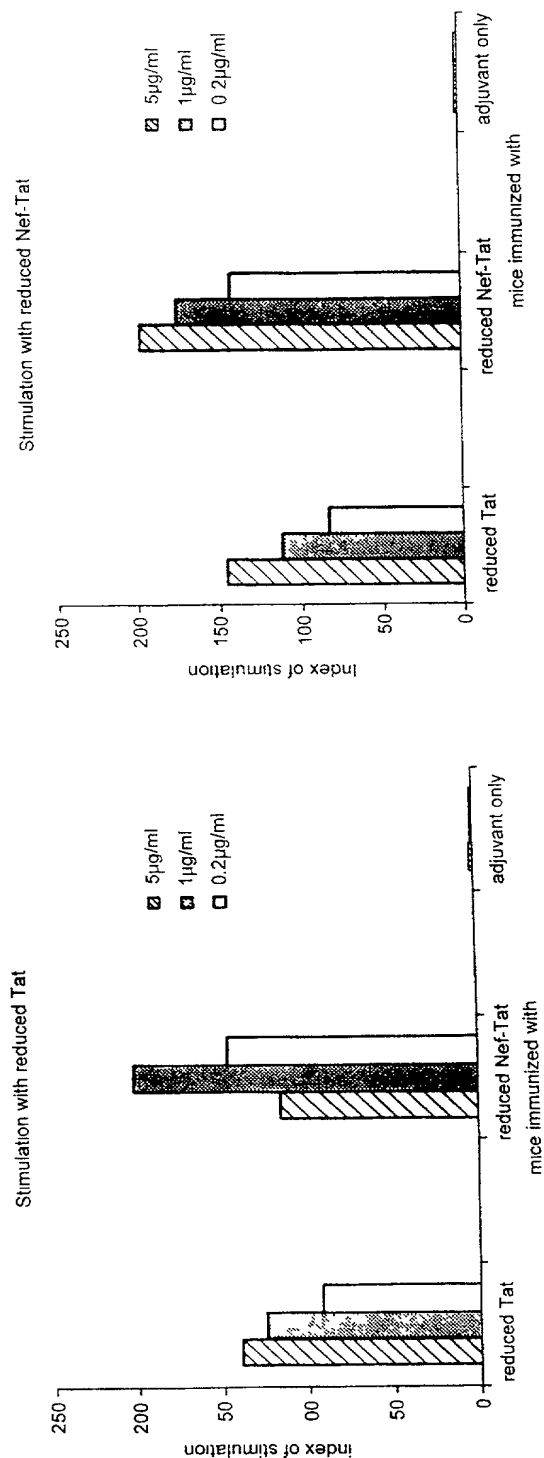
**Fig. 6B**

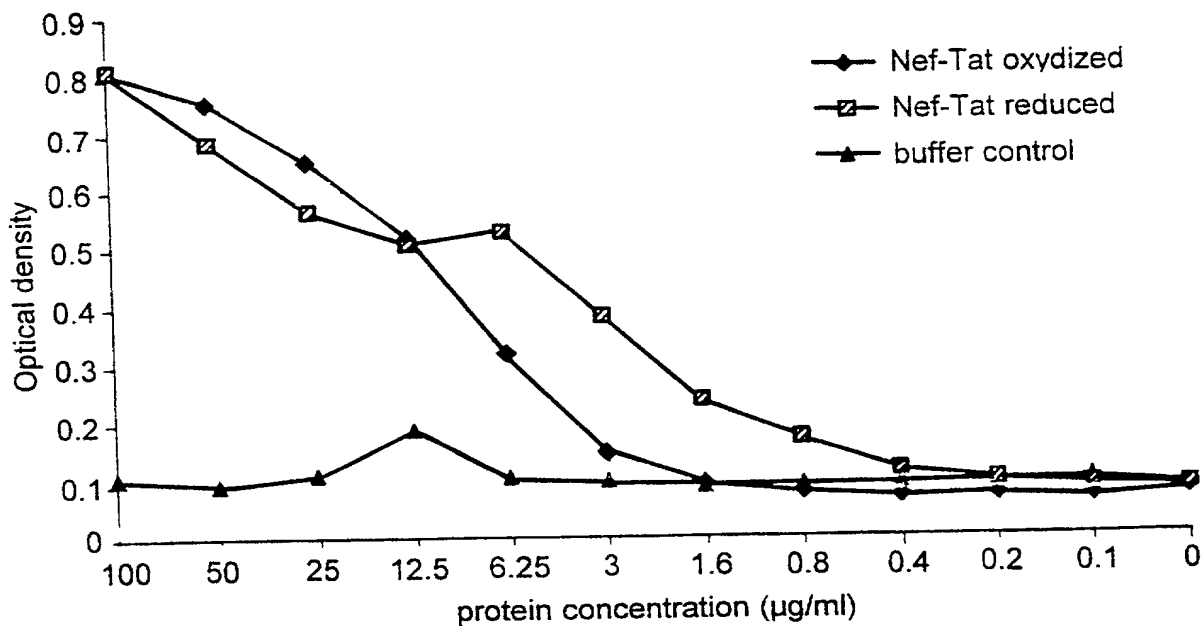
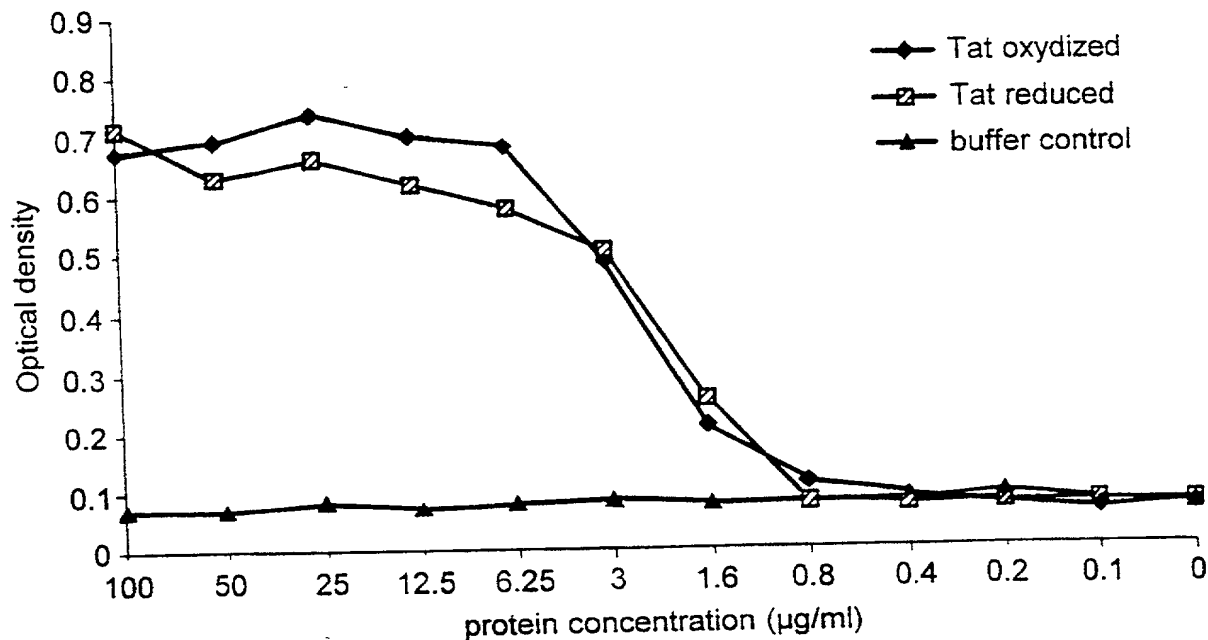
group	immunization	midpoint titers				
		lg	lgG1	lgG2a	lgG2b	ratio lgG1/lgG2a
1	reduced Tat	212799	123242	62697	55763	1,966
2	reduced Nef-Tat	75676	84046	18449	11692	4,556
3	adjuvant only	<4000	<4000	<4000	<4000	

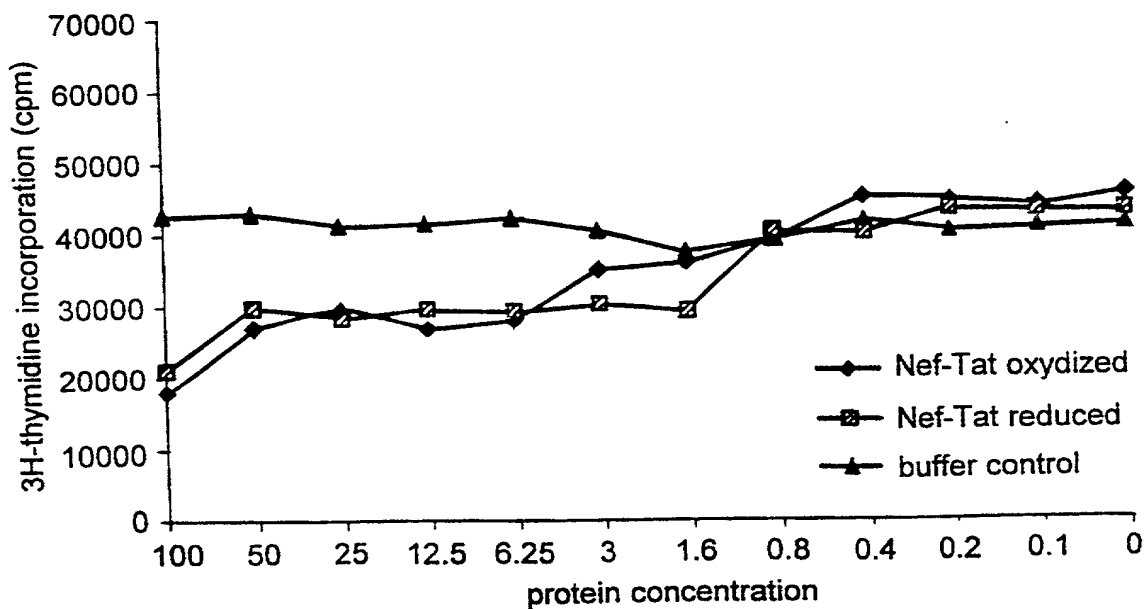
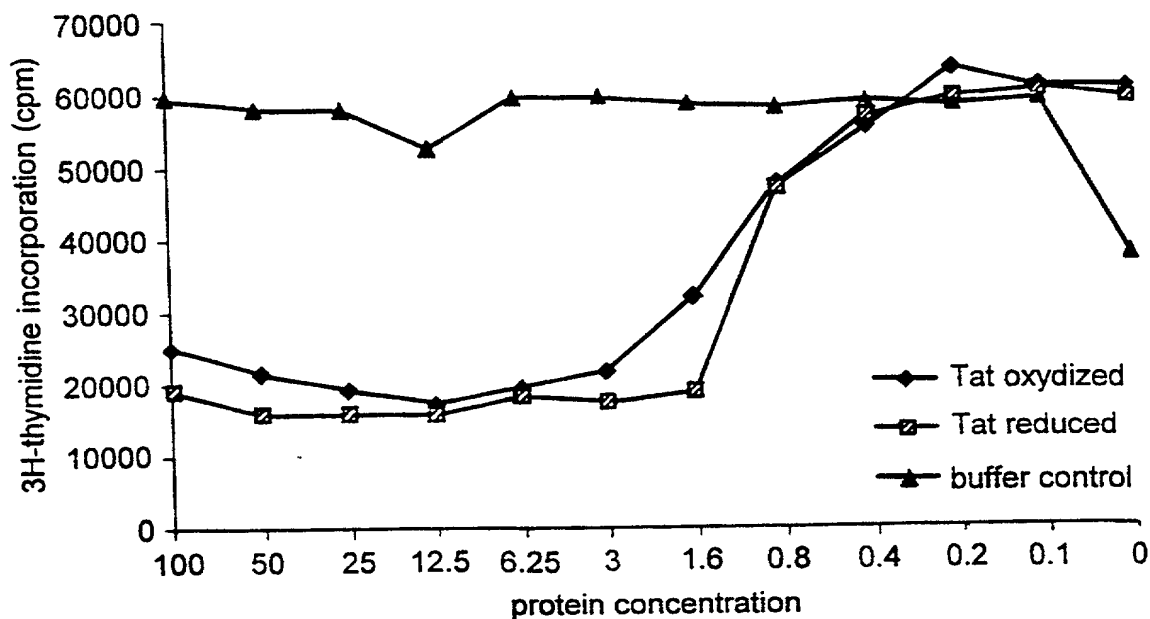


**Fig. 7** Antigen-specific lymphoproliferative response of pooled lymph node cells

[3H] Thymidine incorporation in cpm				Data expressed as stimulation index		
	Group 1 reduced Tat	Group 2 reduced Nef-Tat	Group 3 adjuvant only	Group 1 reduced Tat	Group 2 reduced Nef-Tat	Group 3 adjuvant only
reduced Tat						
5µg/ml	41967	18511	789	140	115	1
1µg/ml	37609	32346	415	125	201	1
0.2µg/ml	27840	23408	397	92	145	1
reduced Nef-Tat						
5µg/ml	43882	31694	483	146	197	1
1µg/ml	33865	28094	245	113	174	0
0.2µg/ml	25079	22891	383	84	142	1
medium	300	161	571	1	1	1



**Fig. 8** Cell binding assay

**Fig. 9** Inhibition of cell growth

## DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Fusion Proteins Comprising HIV-1 Tat and/or Nef Proteins

the specification of which (check one)

☐ is attached hereto.

☒ was filed on 17 September 1998 as Serial No. PCT/EP98/06040  
and was amended on (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below any foreign application for patent or Inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Number	Country	Filing Date	Priority Claimed
9720585.0	Great Britain	26 September 1997	Yes

I hereby claim the benefit under Title 35, United States Code, Section 119(e) of any United States provisional application(s) listed below.

Application Number	Filing Date
--------------------	-------------

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

Serial No.	Filing Date	Status
------------	-------------	--------

2-60 000000-000000 1-60

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full Name of Inventor: Claudine BRUCK

Inventor's Signature: CBruce Date: 7 March 2000  
Residence: Rixensart, Belgium BEX

Residence: Rixensart, Belgium

Citizenship: BELGIAN

Post Office Address: **SmithKline Beecham Corporation**  
**Corporate Intellectual Property - UW2220**  
**P.O. Box 1539**  
**King of Prussia, Pennsylvania 19406-0939**

Full Name of Inventor: Stephane Andre Georges GODART

Inventor's Signature: [Signature] Date: 03 March 2000

Residence: Rixensart, Belgium

Citizenship: BELGIAN

Post Office Address: SmithKline Beecham Corporation  
Corporate Intellectual Property - UW2220  
P.O. Box 1539  
King of Prussia, Pennsylvania 19406-0939



Full Name of Inventor: Martine MARCHAND

Inventor's Signature: [Signature] Date: 07/20/2007

Residence: Glabais, Belgium

Citizenship: BELGIAN

Post Office Address: SmithKline Beecham Corporation  
Corporate Intellectual Property - UW2220  
P.O. Box 1539  
King of Prussia, Pennsylvania 19406-0939

Figure 1 shows the results of the first two experiments. The first experiment was designed to test the hypothesis that the effect of the number of items on the number of items recalled is a function of the number of items presented. The results of this experiment are shown in Figure 1a. The number of items recalled increases with the number of items presented, and the effect of the number of items presented is significant ( $F(4, 12) = 10.0, p < 0.01$ ). The second experiment was designed to test the hypothesis that the effect of the number of items on the number of items recalled is a function of the number of items presented. The results of this experiment are shown in Figure 1b. The number of items recalled increases with the number of items presented, and the effect of the number of items presented is significant ( $F(4, 12) = 10.0, p < 0.01$ ).